Synthesis and in-depth analysis of natural and engineered cystine-knot miniproteins

TECHNISCHE UNIVERSITÄT DARMSTADT

Vom Fachbereich Chemie der Technischen Universität Darmstadt

zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigte kumulative Dissertation,
vorgelegt von

Dipl.-Ing. Michael Reinwarth
aus Langen

Referent:	Prof. Dr. Harald Kolmar
Korreferentin:	Prof. Dr. Katja Schmitz

Tag der Einreichung:	20. Mai 2016
Tag der mündlichen Prüfung:	11. Juli 2016

Darmstadt 2017

D 17

Die vorliegende Arbeit wurde unter der Leitung von Herrn Prof. Dr. Harald Kolmar am Clemens Schöpf-Institut für Organische Chemie und Biochemie der Technischen Universität Darmstadt seit Oktober 2009 angefertigt.

Das Unmögliche möglich zu machen wird ein Ding der Unmöglichkeit.
Andreas Brehme

Publikationsliste

Die Ergebnisse, die dieser Arbeit zu Grunde liegen wurden wie folgt publiziert:

1. Reinwarth, M.; Nasu, D.; Kolmar, H.; Avrutina, O., Chemical synthesis, backbone cyclization and oxidative folding of cystine-knot peptides: promising scaffolds for applications in drug design. *Molecules* **2012,** 17, (11), 12533-52.
2. Reinwarth, M.; Glotzbach, B.; Tomaszowski, M.; Fabritz, S.; Avrutina, O.; Kolmar, H., Oxidative folding of peptides with cystine-knot architectures: kinetic studies and optimization of folding conditions. *Chembiochem* **2013,** 14, (1), 137-46.
3. Glotzbach, B.; Reinwarth, M.; Weber, N.; Fabritz, S.; Tomaszowski, M.; Fittler, H.; Christmann, A.; Avrutina, O.; Kolmar, H., Combinatorial Optimization of Cystine-Knot Peptides towards High-Affinity Inhibitors of Human Matriptase-1. *PLoS One* **2013,** 8, (10), e76956.
4. Reinwarth, M.; Avrutina, O.; Fabritz, S.; Kolmar, H., Fragmentation follows structure: top-down mass spectrometry elucidates the topology of engineered cystine-knot miniproteins. *PLoS One* **2014,** 9, (10), e108626.

An folgenden Publikationen wurde in Form von Koautorenschaften mitgewirkt:

5. Empting, M.; Avrutina, O.; Meusinger, R.; Fabritz, S.; Reinwarth, M.; Biesalski, M.; Voigt, S.; Buntkowsky, G.; Kolmar, H., "Triazole bridge": disulfide-bond replacement by ruthenium-catalyzed formation of 1,5-disubstituted 1,2,3-triazoles. *Angew Chem Int Ed Engl* **2011,** 50, (22), 5207-11.
6. Fabritz, S.; Horner, S.; Konning, D.; Empting, M.; Reinwarth, M.; Dietz, C.; Glotzbach, B.; Frauendorf, H.; Kolmar, H.; Avrutina, O., From pico to nano: biofunctionalization of cube-octameric silsesquioxanes by peptides and miniproteins. *Org Biomol Chem* **2012,** 10, (31), 6287-93.
7. Glotzbach, B.; Schmelz, S.; Reinwarth, M.; Christmann, A.; Heinz, D. W.; Kolmar, H., Structural characterization of Spinacia oleracea trypsin inhibitor III (SOTI-III). *Acta Crystallogr D Biol Crystallogr* **2013,** 69, (Pt 1), 114-20.
8. Maass, F.; Wustehube-Lausch, J.; Dickgiesser, S.; Valldorf, B.; Reinwarth, M.; Schmoldt, H. U.; Daneschdar, M.; Avrutina, O.; Sahin, U.; Kolmar, H., Cystine-knot peptides targeting cancer-relevant human cytotoxic T lymphocyte-associated antigen 4 (CTLA-4). *J Pept Sci* **2015,** 21, (8), 651-60.

Danksagung

Zu allererst möchte ich denen danken, ohne die die Entstehung dieser Arbeit unmöglich gewesen wäre:

Katrin Reinwarth: Danke dafür, dass ich dich meine Frau nennen darf. Danke, dass du immer zu mir stehst – auch in schlechten Zeiten. Danke, dass du mich komplett machst. Danke für die vergangenen Jahre – und danke für die kommenden. Ich freue mich drauf!

Lukas und Amelie Reinwarth: Danke, dass für die Abwechslung und Ablenkung von der Arbeit. Danke dafür, dass ihr mich mit eurer unbeschwerten Lebensfreude zum Lachen bringt, wie nur ein Kind es kann. Behaltet eure ungetrübte Lebensfreude solange ihr könnt und lasst mich daran teilhaben!

Prof. Harald Kolmar: Danke für die Möglichkeit diese Arbeit unter deiner Regie zu erstellen! Danke für deine Hilfe, die die Fokussierung beibehält, aber trotzdem genug Platz für eigene Ideen lässt. Danke für die konstruktive Kritik, die immer hilft sich zu verbessern und das Lob, das einen motiviert weiterzumachen und nicht aufzugeben. Danke für einen Chef, der immer da ist, wenn man ihn braucht!

Dr. Olga Avrutina: Danke für die schnelle und unkomplizierte Hilfe, falls es mal nicht weiterging. Danke, dass du deine praktischen Erfahrungen an mich weitergeben hast. Danke, dass du mich sorgfältiges und wissenschaftliches Arbeiten gelehrt hast. Und natürlich danke für die Hilfe beim Schreiben der Publikationen – bei dir habe ich mehr Englisch gelernt als in der Schule.

Inge und Manfred Grentz: Danke, dass ihr immer für uns da wart und wir uns bei euch wohl und willkommen fühlen durften.

Die gesamte Arbeitsgruppe Kolmar: Ich möchte allen für die tolle Zusammenarbeit danken für die Hilfestellung und Lösungsvorschläge bei meinen Problemen und natürlich für die spannenden ‚crossover-Projekte'. Zwei Einzelpersonen gehört besonderen Dank: **Dr. Bernhard Glotzbach** und **Dr.-Ing. Sebastian Fabritz**: Danke für die Entwicklung und Durchführung unserer gemeinsamen Projekte!

Den Verlagen **Wiley VCH**, **MDPI** und **PLOS ONE** für die Erlaubnis die nachfolgenden Publikationen in meiner Dissertation zu verwenden und zu verwenden.

Inhaltsverzeichnis

1 Überblick über das Forschungsfeld und Zielsetzung der Arbeit

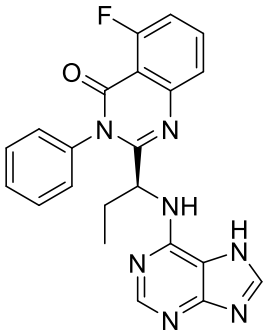
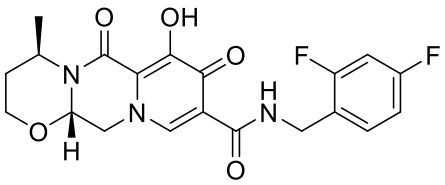
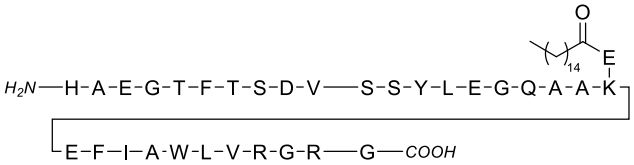
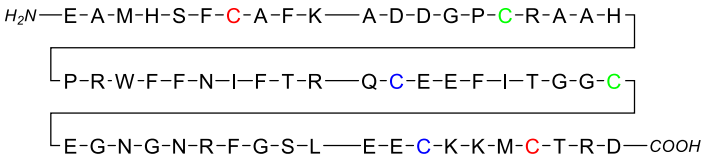
Die fortwährende Weiterentwicklung der individualisierten Medizin stellt hohe Anforderungen an neu entwickelte Medizinprodukte. Im Gegensatz zur Behandlung oder Erkennung von Krankheiten, die viralen oder bakteriellen Ursprungs sind, ist die Adressierung von körpereigenen Zellen mit Fehlfunktionen wie Alzheimer und Malignomen ungleich schwieriger. Hier muss das verabreichte Therapeutikum oder Diagnostikum hinreichend selektiv wirken, um eine Behandlung oder Markierung vitaler Körperzellen zu vermeiden. Die Kenntnis der veränderten Beschaffenheit der Zelloberfläche ist dabei ebenso unabdingbar wie die der Struktur und Wirkmechanismen der zellbindenden Präparate.

Die Entwicklung neuer Therapeutika oder Diagnostika basiert heutzutage in erster Linie auf drei Substanzklassen: niedermolekulare Verbindungen (engl. *Small molecules*), Peptide und Antikörper (Tabelle 1).

Nach der Identifikation einer potentiell aktiven Substanzklasse erfolgt die Synthese von *small molecules* aufgrund der konstanten Gerüststruktur oftmals durch kombinatorische Hochdurchsatzverfahren, da hier nur Änderungen von Seitenketten oder funktionellen Gruppen der Peripherie vorgenommen werden.[1] Dabei werden Synthese und chromatografische und massenspektrometrische Analytik gekoppelt. Als Ausgangspunkt dienen gewöhnlich körpereigene Substanzen wie z. B. Substratanaloga, die an das Enzym binden und damit die hydrolytische Aktivität des Enzyms inhibieren.[1] Auf diesem Wege können binnen kürzester Zeit hunderte biologisch hochaktiver Substanzen synthetisiert und auf ihre Wirksamkeit getestet werden. Obwohl auf diesem Wege hochaffine Moleküle mit Bindungskonstanten im subnanomolaren Bereich entwickelt werden können, wird deren Wirkung aber häufig durch eine geringe Selektivität beeinträchtigt, was in eine Schädigung vitaler Zellen resultieren kann. Dies führt letztlich zu einer erhöhten Toxizität dieser Substanzklasse.[2] Ebenso ist die Resorption aufgrund der strukturellen Eigenschaften, wie Lipophilie und fehlender Ladungen oftmals reduziert.[1, 3] Eine Vielzahl der heute gängigen Chemotherapeutika wie Imatinib, *Nexavar®* (Sorafenib) und *Tarceva®* (Erlotinib) auf diesem Wege entwickelt worden.[1]

Antikörper bestehen aus je zwei leichten und schweren Ketten, die wiederum in konstante und variable Bereiche unterteilt sind. Die Bindung zum entsprechenden Antigen erfolgt über die variablen Bereiche. Innerhalb dieser Domänen befinden sich die je drei hypervariablen Regionen, die sogenannten CDRs (*complementary determining region*), die für die Antigen-Antikörper-Bindung verantwortlich sind.[4] Die Bindung einer einzelnen CDR zu ihrem jeweiligen Antigen ist relativ schwach, die Bindungskonstanten liegen hier in der Regel im mikromolaren Bereich.[4] Durch die simultane Bindung aller drei CDRs an das Antigen potenzieren sich die Bindungskonstanten bis in den femtomolaren Bereich. Durch dieses Zusammenspiel der einzelnen CDRs ist die Selektivität im Vergleich zu den small-molecule-basierenden Präparaten stark erhöht.[4] Ein weiterer, größenbedingter (bis zu 1000 kDa) Vorteil der Antikörper ist deren langsamer Abbau im Körper des Patienten, was einen längeren Wirkzeitraum zur Folge hat.[5]

Tabelle 1: Therapeutika verschiedener Substanzklassen. Cystine sind aus Übersichtlichkeit farblich gekennzeichnet.

Name, Markenname	**Struktur**	**Substanz-klasse**	**Zielprotein, bevorzugter Einsatz**
Idelalisib, *Zydelig®*		small-molecule	Kinase (*phosphoinositide 3-kinase*), Leukämie
Dolutegravir, *Tivicay®*		small-molecule	Integrase, HIV-1
Liraglutide, *Victoz®*		Peptid	GLP-1 (*glucagon-like peptide-1)* Rezeptor, Typ 2 diabetes
Ecallantide, *Kalbitor®*		Peptid	Kallikrein, vererbtes Angioödem
Nivolumab, *Opdivo®*	Humaner IG4-Antikörper	Antikörper	PD-1 (*programmed cell death protein 1*), verschiedene Krebsarten
Ramucirumab *Cyramza®*	Humaner IG1-Antikörper	Antikörper	VEGFR 2 (*vascular endothelial growth factor recptor*), Magenkarzinom

Nachteilig wirkt sich hingegen die Größe in Bezug auf die Gewebsdurchdringung aus, was insbesondere bei schlecht durchbluteten Tumorgeweben in einer verminderten Bioverfügbarkeit resultiert.[5] Außerdem sind die Synthese und Reinigung teilweise sehr aufwendig, da die meist glykosylierten Antikörper aus Zellkulturen gewonnen und anschließend chromatografisch gereinigt werden müssen.[6, 7] Trotz dieser Hürden wird eine Vielzahl monoklonaler Antikörper in der Diagnostik als Tumormarker, in der Immunoszintigraphie und als Therapeutika kommerziell eingesetzt.[8]

Peptid-basierende Therapeutika oder Diagnostika vereinen eine vereinfachte Synthese im Vergleich zu Antikörperpräparaten und eine verbesserte Gewebsdurchdringung.[9] Allerdings weisen sie eine verringerte Serumhalbwertszeit und Selektivität bei vergleichbaren Affinitäten auf.[9-11] Diese Eigenschaften machen Peptide zu einer interessanten Alternative zu klassischen Antikörper- oder *small molecule* Präparaten. In den letzten Jahren hat sich der Anteil der durch die amerikanische

Food and Drug Administration (FDA) zugelassen peptid-basierten Therapeutika signifikant erhöht.[12-17]

Eine Molekülklasse, die viele Vorteile aller drei erwähnten Substanzklassen in sich vereint sind die Cystinknotenminiproteine. Diese sind Peptide mit ca. 30 Aminosäuren, die durch drei ‚pseudoknotenbildende' Cystine sowie ein extensives Netzwerk von Wasserstoffbrücken eine stabile dreidimensionale Struktur und eine außerordentliche Stabilität aufweisen (Abbildung 1).[18-24] Diese strukturellen Eigenschaften führen zu einer hohen Stabilität gegenüber thermischen und chemischen Einflüssen sowie proteolytischer Hydrolyse, was eine orale Verabreichung von Therapeutika denkbar macht.[18, 25-27] Unterschieden werden die meist antibiotisch wirkenden, zyklischen und die inhibitorischen Miniproteine.[23, 28] Aufgrund der relativ geringen Größe von üblicherweise bis zu 35 Aminosäuren kann die Synthese sowohl rekombinant als auch chemisch in relativ hohen Ausbeuten und mit vertretbarem finanziellem Aufwand durchgeführt werden.[23, 29, 30]

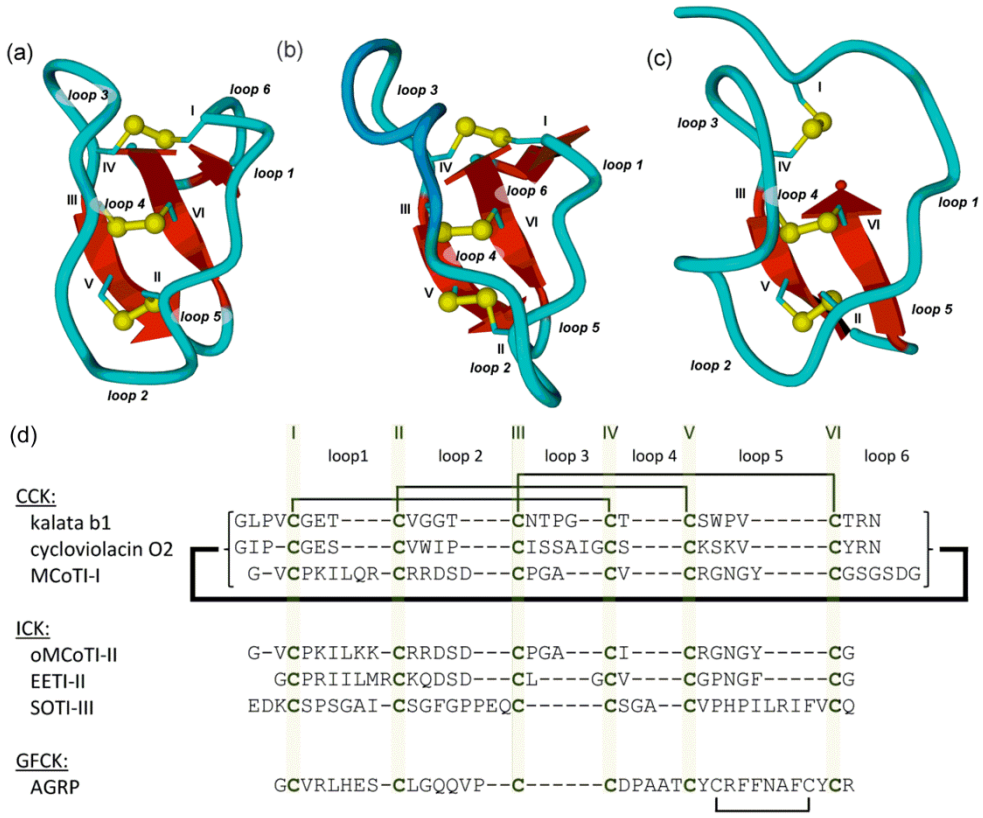


Abbildung 1: Cartoondiagramme und Aminosäuresequenzen typischer Cystinknoten. (a-c):Schleifen sind in hellblau dargestellt und entsprechend ihrem Vorkommen innerhalb der Sequenz nummeriert, α-Helices in dunkelblau, β-Faltblätter in rot und Cysteine in gelb mit römischer Nummerierung entsprechend ihrem Vorkommen innerhalb der Peptidsequenz. (a) *Möbius* Cyclotid Kalata B1. (PDB-ID: 1NB1) (b) *Bracelet* Cyclotid Cycloviolacin O2. (PDB-ID: 2KNM) (c) Azyklischer Inhibitor cystine knot ocMCoTI-II (PDB-ID: 2IT8). Strukturen sind mit Yasara Ver. 12.4.1 dargestellt. (d) Cysteine sind entsprechend ihrem Vorkommen innerhalb der Sequenz nummeriert, Cystinverbindungen sowie die Makrozyklisierung sind schematisch dargestellt.[23]

Des Weiteren können über eine Änderung der bindenden Schleife (Abbildung 1) Affinitäten und Selektivitäten von inhibitorischen Cystinknoten-Miniproteinen verändert werden. Durch eine

Randomisierung dieser Schleife wurden aus zufallsmäßig oder gezielt erzeugten Bibliotheken hochaffine und -selektive Inhibitoren gegen krebsrelevante Enzyme gefunden.[31, 32] Ebenso konnten durch rationale Veränderungen dieser Schleife Binder für verschiedene Tumormarker generiert werden.[21, 33-36]

In der hier vorgelegten Arbeit sollte die Synthese und Analytik bekannter und in ihrer Sequenz veränderter Varianten offenkettiger, inhibitorischer Miniproteine bearbeitet werden:

- Die Cysteinoxidation und die damit verbundene Ausbildung sekundärer und tertiärer Strukturelemente von Miniproteinen ist ein entscheidender Schritt der Synthese.[23] Die Optimierung der Ausbeute der peptidchemischen Darstellung der Peptidkette war daher ebenso von besonderem Interesse wie die Verbesserung der synthetischen Bedingungen und Ausbeute bei der Ausbildung des Cystinknotens. Diese Optimierung war insbesondere bei Miniproteinen mit veränderter Primärstruktur notwendig, da die bisher verwendeten Systeme zur oxidativen Faltung häufig sehr kompliziert sind, nur für eine einzelne Familie von Miniproteinen anwendbar ist oder ur geringe Ausbeuten liefert.

- Daraus resultierend sollte im Anschluss die physikochemische Charakterisierung der Miniproteine durchgeführt werden. Mit Hilfe dieser Technologien sollten neue Erkenntnisse über die Primär- Sekundär- und Tertiärstruktur gewonnen werden. Insbesondere die Verifizierung der korrekten Disulfidtopologie ist hier von besonderem wissenschaftlichem Interesse und nicht trivial.

- Die aus der Strukturaufklärung erhaltenen Erkenntnisse sollten dann wiederum zur Generierung von Peptidbibliotheken verwendet werden, deren Durchmusterung zum Erhalt hochaffiner und –selektiver Binder krebsrelevanter Oberflächenenzyme führen sollte. Diese sollten dann mittels der optimierten Synthesebedingungen hergestellt werden und mittels enzymologischer Tests biochemisch charakterisiert werden. Die hinsichtlich Affinität und Selektivität potentesten Inhibitoren sollten dann in Dosis-Wirkungs-Studien auf Prostatakrebszellen *ex vivo* hinsichtlich ihrer inhibitorischen Aktivität hinsichtlich humaner Matriptase-1 charakterisiert werden.

Referenzen

1. Hoelder, S.; Clarke, P. A.; Workman, P., Discovery of small molecule cancer drugs: successes, challenges and opportunities. *Mol Oncol* **2012,** 6, (2), 155-76.
2. Toporkiewicz, M.; Meissner, J.; Matusewicz, L.; Czogalla, A.; Sikorski, A. F., Toward a magic or imaginary bullet? Ligands for drug targeting to cancer cells: principles, hopes, and challenges. *Int J Nanomedicine* **2015,** 10, 1399-414.
3. Prueksaritanont, T.; Tang, C., ADME of biologics-what have we learned from small molecules? *AAPS J* **2012,** 14, (3), 410-9.
4. Weiner, G. J., Building better monoclonal antibody-based therapeutics. *Nat Rev Cancer* **2015,** 15, (6), 361-70.
5. Zhao, L.; Ren, T. H.; Wang, D. D., Clinical pharmacology considerations in biologics development. *Acta Pharmacol Sin* **2012,** 33, (11), 1339-47.

6. Picanco-Castro, V.; de Freitas, M. C.; Bomfim Ade, S.; de Sousa Russo, E. M., Patents in therapeutic recombinant protein production using mammalian cells. *Recent Pat Biotechnol* **2014,** 8, (2), 165-71.
7. Dorai, H.; Ganguly, S., Mammalian cell-produced therapeutic proteins: heterogeneity derived from protein degradation. *Curr Opin Biotechnol* **2014,** 30, 198-204.
8. Guan, M.; Zhou, Y. P.; Sun, J. L.; Chen, S. C., Adverse events of monoclonal antibodies used for cancer therapy. *Biomed Res Int* **2015,** 2015, 428169.
9. Rosenmayr-Templeton, L., The oral delivery of peptides and proteins: established versus recently patented approaches. *Pharm Pat Anal* **2013,** 2, (1), 125-45.
10. Vlieghe, P.; Lisowski, V.; Martinez, J.; Khrestchatisky, M., Synthetic therapeutic peptides: science and market. *Drug Discov Today* **2010,** 15, (1-2), 40-56.
11. Bray, B. L., Large-scale manufacture of peptide therapeutics by chemical synthesis. *Nat Rev Drug Discov* **2003,** 2, (7), 587-93.
12. Mullard, A., 2014 FDA drug approvals. *Nat Rev Drug Discov* **2015,** 14, (2), 77-81.
13. Mullard, A., 2013 FDA drug approvals. *Nat Rev Drug Discov* **2014,** 13, (2), 85-9.
14. Mullard, A., 2012 FDA drug approvals. *Nat Rev Drug Discov* **2013,** 12, (2), 87-90.
15. Mullard, A., 2011 FDA drug approvals. *Nat Rev Drug Discov* **2012,** 11, (2), 91-4.
16. Mullard, A., 2010 FDA drug approvals. *Nat Rev Drug Discov* **2011,** 10, (2), 82-5.
17. Albericio, F.; Kruger, H. G., Therapeutic peptides. *Future Med Chem* **2012,** 4, (12), 1527-31.
18. Heitz, A.; Avrutina, O.; Le-Nguyen, D.; Diederichsen, U.; Hernandez, J. F.; Gracy, J.; Kolmar, H.; Chiche, L., Knottin cyclization: impact on structure and dynamics. *BMC Struct Biol* **2008,** 8, 54.
19. Avrutina, O.; Schmoldt, H. U.; Gabrijelcic-Geiger, D.; Wentzel, A.; Frauendorf, H.; Sommerhoff, C. P.; Diederichsen, U.; Kolmar, H., Head-to-tail cyclized cystine-knot peptides by a combined recombinant and chemical route of synthesis. *Chembiochem* **2008,** 9, (1), 33-7.
20. Kratzner, R.; Debreczeni, J. E.; Pape, T.; Schneider, T. R.; Wentzel, A.; Kolmar, H.; Sheldrick, G. M.; Uson, I., Structure of Ecballium elaterium trypsin inhibitor II (EETI-II): a rigid molecular scaffold. *Acta Crystallogr D Biol Crystallogr* **2005,** 61, (Pt 9), 1255-62.
21. Ackerman, S. E.; Currier, N. V.; Bergen, J. M.; Cochran, J. R., Cystine-knot peptides: emerging tools for cancer imaging and therapy. *Expert Rev Proteomics* **2014,** 11, (5), 561-72.
22. Gracy, J.; Chiche, L., Structure and modeling of knottins, a promising molecular scaffold for drug discovery. *Curr Pharm Des* **2011,** 17, (38), 4337-50.
23. Reinwarth, M.; Nasu, D.; Kolmar, H.; Avrutina, O., Chemical synthesis, backbone cyclization and oxidative folding of cystine-knot peptides: promising scaffolds for applications in drug design. *Molecules* **2012,** 17, (11), 12533-52.
24. Rosengren, K. J.; Daly, N. L.; Plan, M. R.; Waine, C.; Craik, D. J., Twists, knots, and rings in proteins. Structural definition of the cyclotide framework. *J Biol Chem* **2003,** 278, (10), 8606-16.
25. Kolmar, H., Natural and engineered cystine knot miniproteins for diagnostic and therapeutic applications. *Curr Pharm Des* **2011,** 17, (38), 4329-36.
26. Kolmar, H., Biological diversity and therapeutic potential of natural and engineered cystine knot miniproteins. *Curr Opin Pharmacol* **2009,** 9, (5), 608-14.
27. Gould, A.; Ji, Y.; Aboye, T. L.; Camarero, J. A., Cyclotides, a novel ultrastable polypeptide scaffold for drug discovery. *Curr Pharm Des* **2011,** 17, (38), 4294-307.
28. Smith, A. B.; Daly, N. L.; Craik, D. J., Cyclotides: a patent review. *Expert Opin Ther Pat* **2011,** 21, (11), 1657-72.
29. Reinwarth, M.; Glotzbach, B.; Tomaszowski, M.; Fabritz, S.; Avrutina, O.; Kolmar, H., Oxidative folding of peptides with cystine-knot architectures: kinetic studies and optimization of folding conditions. *Chembiochem* **2013,** 14, (1), 137-46.

30. Hofmeyer, T.; Bulani, S. I.; Grzeschik, J.; Krah, S.; Glotzbach, B.; Uth, C.; Avrutina, O.; Brecht, M.; Goringer, H. U.; van Zyl, P.; Kolmar, H., Protein production in Yarrowia lipolytica via fusion to the secreted lipase Lip2p. *Mol Biotechnol* **2014,** 56, (1), 79-90.
31. Glotzbach, B.; Reinwarth, M.; Weber, N.; Fabritz, S.; Tomaszowski, M.; Fittler, H.; Christmann, A.; Avrutina, O.; Kolmar, H., Combinatorial Optimization of Cystine-Knot Peptides towards High-Affinity Inhibitors of Human Matriptase-1. *PLoS One* **2013,** 8, (10), e76956.
32. Maass, F.; Wustehube-Lausch, J.; Dickgiesser, S.; Valldorf, B.; Reinwarth, M.; Schmoldt, H. U.; Daneschdar, M.; Avrutina, O.; Sahin, U.; Kolmar, H., Cystine-knot peptides targeting cancer-relevant human cytotoxic T lymphocyte-associated antigen 4 (CTLA-4). *J Pept Sci* **2015,** 21, (8), 651-60.
33. Moore, S. J.; Leung, C. L.; Norton, H. K.; Cochran, J. R., Engineering agatoxin, a cystine-knot peptide from spider venom, as a molecular probe for in vivo tumor imaging. *PLoS One* **2013,** 8, (4), e60498.
34. Silverman, A. P.; Kariolis, M. S.; Cochran, J. R., Cystine-knot peptides engineered with specificities for alpha(IIb)beta(3) or alpha(IIb)beta(3) and alpha(v)beta(3) integrins are potent inhibitors of platelet aggregation. *J Mol Recognit* **2011,** 24, (1), 127-35.
35. Silverman, A. P.; Levin, A. M.; Lahti, J. L.; Cochran, J. R., Engineered cystine-knot peptides that bind alpha(v)beta(3) integrin with antibody-like affinities. *J Mol Biol* **2009,** 385, (4), 1064-75.
36. Ji, Y.; Majumder, S.; Millard, M.; Borra, R.; Bi, T.; Elnagar, A. Y.; Neamati, N.; Shekhtman, A.; Camarero, J. A., In vivo activation of the p53 tumor suppressor pathway by an engineered cyclotide. *J Am Chem Soc* **2013,** 135, (31), 11623-33.

2 Einleitung

2.1 Chemical Synthesis, Backbone Cyclization and Oxidative Folding of Cystine-knot Peptides – Promising Scaffolds for applications in Drug Design

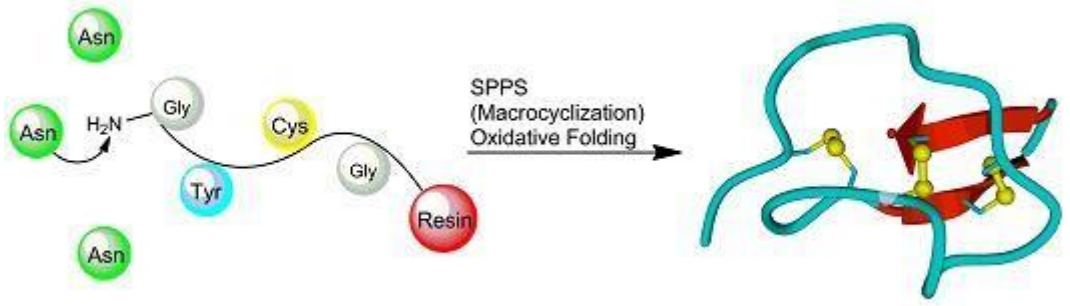
Title

Chemical Synthesis, Backbone Cyclization and Oxidative Folding of Cystine-knot Peptides — Promising Scaffolds for Applications in Drug Design

Authors

Michael Reinwarth [†], Daichi Nasu [†], Harald Kolmar and Olga Avrutina *

Graphical Abstract



Bibliographical Data

Molecules **2012**, *17*(11), 12533-12552 (This article belongs to the Special Issue Chemical Protein and Peptide Synthesis)

doi:10.3390/molecules171112533

Received: 15 August 2012 / Revised: 19 September 2012 / Accepted: 22 October 2012 / Published: 24 October 2012

Contributions by Michael Reinwarth

- Performed literature research
- Wrote manuscript and created tables
- Design of Figures (together with Daichi Nasu)

.

Molecules **2012**, *17*, 12533-12552; doi:10.3390/molecules171112533

OPEN ACCESS

molecules

ISSN 1420-3049

www.mdpi.com/journal/molecules

Review

Chemical Synthesis, Backbone Cyclization and Oxidative Folding of Cystine-knot Peptides — Promising Scaffolds for Applications in Drug Design

Michael Reinwarth [†], Daichi Nasu [†], Harald Kolmar and Olga Avrutina *

Institute for Organic Chemistry and Biochemistry, Technische Universität Darmstadt, Petersenstraße 22, D-64287 Darmstadt, Germany

[†] These authors contributed equally to this work.

* Author to whom correspondence should be addressed; E-Mail: Avrutina@Biochemie-TUD.de; Tel.: +49-6151-164758; Fax: +49-6151-165399.

Received: 15 August 2012; in revised form: 19 September 2012 / Accepted: 22 October 2012 / Published: 24 October 2012

Abstract: Cystine-knot peptides display exceptional structural, thermal, and biological stability. Their eponymous motif consists of six cysteine residues that form three disulfide bonds, resulting in a notably rigid structural core. Since they highly tolerate either rational or combinatorial changes in their primary structure, cystine knots are considered to be promising frameworks for the development of peptide-based pharmaceuticals. Despite their relatively small size (two to three dozens amino acid residues), the chemical synthesis route is challenging since it involves critical steps such as head-to-tail cyclization and oxidative folding towards the respective bioactive isomer. Herein we describe the topology of cystine-knot peptides, their synthetic availability and briefly discuss potential applications of engineered variants in diagnostics and therapy.

Keywords: CCK; cyclotide; cystine knot; ICK; inhibitor; knottin; miniprotein; native chemical ligation; oxidative folding

1. Introduction

Cystine-knot peptides, also termed knottins, are promising scaffolds for the design of peptide-based pharmaceuticals as they combine potent bioactivities with remarkable thermal and proteolytic

stabilities [1–3]. Their amide backbone of approximately 30 amino acid residues is compacted by three disulfide bonds forming a characteristic 'pseudo-knotted' structure [4]. Cystine-knot peptides can be divided into three major subclasses: inhibitor cystine knots (ICK), cyclic cystine knots (CCK) and growth factor cystine knots (GFCK). While the majority of cystine-knot protease inhibitors comprises a linear backbone and displays inhibition constants in the low nanomolar to picomolar range, CCK peptides are defined by a head-to-tail backbone cyclization motif [5]. Compared to ICK and CCK, GFCK peptides are larger, less stable and commonly produced recombinantly. Hence, they will not be considered here.

The systematic research in the field of cystine-knot peptides arose in the early 1970s when a cyclotide kalata B1 was identified in Congo, where women brew tea from the leaves of the plant *Oldenlandia affinis* to accelerate childbirth [6]. Nevertheless, it took until the 1990s to finally solve the structural properties of CCK and a short period later also of ICK peptides [7,8].

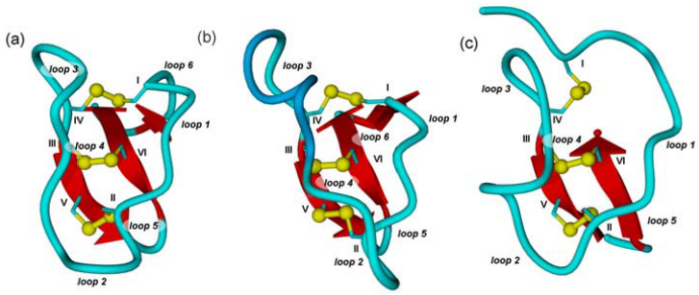
To date, cystine-knot peptides have been found in diverse organisms: arthropoda, fungi, mollusca, plantae, porifera, and vertebrata [9,10]. Due to their wide-spread occurrence in combination with a structurally conserved core and an astonishing diversity with respect to amino acid sequence and function, they can be considered as one of Nature's combinatorial libraries [11].

2. Structure

2.1. The Cystine-Knot Motif

Cystine knots share common structural motifs that are defined by three antiparallel β-strands which are connected through short loops along with a considerable network of hydrogen bonds, and the eponymous knotted disulfide connections [12]. The constrained conformation mainly results from the disulfide bond between CysIII and CysVI (cysteines within the sequence are numbered according to their appearance from the amino- to the carboxy-terminus) which is threaded through the embedded ring formed upon the disulfide linkage of CysI and CysIV as well as CysII and CysV (Figure 1) [13].

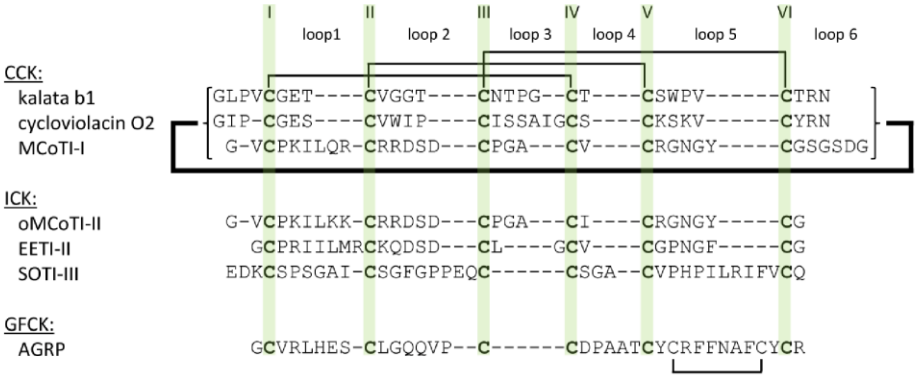
Figure 1. Cartoon diagrams of prototypical cystine knots. Loops are depicted in light blue and numbered according to their appearance in the sequence, α-helices in dark blue, β-sheets in red, and cysteines in yellow with Roman numerals according to their appearance in the sequence. (**a**) Möbius cyclotide kalata B1. (PDB-ID: 1NB1) (**b**) Bracelet cyclotide cycloviolacin O2. (PDB-ID: 2KNM) (**c**) Acyclic inhibitor cystine knot ocMCoTI-II (PDB-ID: 2IT8). Structures modeled with Yasara Ver. 12.4.1.



Despite sequential and numerical differences within the variable loops of the different families, they all share the knotted core merging them into the cystine-knot family. These cystine motifs, in fact, are more important for peptide stability and rigidity than an eventual backbone cyclization [14]. Additionally, an extensive network of hydrogen bonds, especially *via* the β-sheets contributes an essential energetic value to the thermodynamic stability of cystine knots [11,12].

These structural constraints leave the loops (Figure 2) in a surface-exposed state regardless of the hydrophobicity of the assembled amino acid residues. Therefore, also highly hydrophobic residues can be presented on the outer shell of the miniprotein targeting hydrophobic binding pockets or disrupting cell membranes [5,12].

Figure 2. Sequence alignment of certain cystine knots. Cystine connections as well as head-to-tail macrocyclization motif are indicated.



2.2. Cyclic Cystine Knots

Cyclic cystine-knot peptides combine a macrolactam backbone with the knotted disulfide pattern [11]. They are supposed to play an important role in plant defense, as most of them have insecticidal activity due to their ability to disrupt cell membranes [15]. Interestingly, recent studies report antimicrobial, anti-HIV, and cytotoxic activities as well [16–20].

Structurally, cyclotides are divided into Möbius, bracelet, and trypsin inhibitor subclasses. In comparison to other cyclotide families cyclic trypsin inhibitors MCoTI-I and –II extracted from the seeds of *Momordica cochinchinensis* display considerable structural differences in their loop regions, obviously demonstrating similarities with knottins from the squash inhibitor family (Figure 2) [21–23]. Therefore, we share the opinion to categorize them into the ICK family [11,21].

The cyclotide kalata B1 is the prototypic cyclotide of a Möbius type, while cycloviolacin O2 is a common example for a bracelet cyclotide (Figures 1 and 2) [11]. Conformational differences between Möbius and bracelet cyclotides are caused by the presence or absence of a *cis*-proline in loop 5. This moiety induces a twist in the orientation of the central *β*-sheet of Möbius CCK peptides, thus causing their oblate shape compared to bracelet cyclotides which lack this conformationally determinative unit [5].

2.3. Inhibitor Cystine Knots

ICK peptides, also referred to as knottins, are found in the seeds of various plants, among them bitter gourd *Momordica cochinchinensis* (MCoTI I-III), squirting cucumber *Ecballium elaterium* (EETI I-III), and spinach *Spinacia oleracea* (SOTI I-III) (Figures 1 and 2). Their potent inhibitory effect against one of the major digestive proteases, trypsin, indicates their role in zoochory. While MCoTI and EETI are members of the squash inhibitor family with the inhibitory loop located between CysI and CysII, SOTI miniproteins display similarity to a class of antimicrobial peptides from the seeds of *M. jalapa* with CysV and CysVI flanking the inhibitory loop [8,21,24–26].

ICK peptides do not necessarily possess a cyclic backbone (indeed, only MCoTI-I and MCoTI-II are macrocyclic) but are defined according to their inhibitory effect against their respective target proteases. Cyclic trypsin inhibitors have been reported to be more potent than their open-chain counterparts. Nevertheless, inhibition constants of open-chain variants are still in a low nanomolar range [23]. Surprisingly, backbone cyclization only has minor effects on thermal and proteolytic stability providing evidence that the cystine knot motif is mainly responsible for the remarkable robustness of this scaffold [27].

3. Synthesis of Cystine-Knot Peptides

In this section we will critically discuss recombinant and chemical synthesis of cystine-knot peptides. Although the permanently increasing arsenal of reagents, methodologies, and instruments for solid phase peptide synthesis (SPPS) has largely eliminated synthetic problems during chain assembly, backbone cyclization and oxidative folding towards the regioselective formation of multiple disulfide bonds are still the crucial steps during total chemical synthesis of cystine knots and will therefore be considered in further detail [28–33].

3.1. Recombinant Production

In Nature, biosynthesis of cystine-knot peptides has been evolutionary optimized towards high yields of the bioactive forms [34]. Therefore, extraction of cystine-knot peptides from the corresponding plant sources is a common way to isolate the wild type sequences [31,35–37].

In contrast to chemical synthesis, backbone cyclization (rather than oxidative folding) is the crucial step during recombinant production of cyclic cystine-knot peptides as only a few enzymes are known to catalyze the desired amide bond formation [38]. For cyclotides, intein-based cyclization methods have been recently reported [39–41]. Through several enzyme-catalyzed steps that include an S-N acyl shift, intein fragments are finally cleaved off and the peptide termini become condensed *via* a native amide bond (Figure 3) [39–41]. However, intein-mediated cyclization often lacks satisfactory yields and it remains a challenge to obtain multimilligram amounts of the respective cyclic peptide (Table 1). The recently reported production of cyclic MCoTI-I in a BL21 strain of *E. coli* indicates that high-yield synthesis is possible under fine-tuned conditions. [39–42].

Figure 3. Common strategies for backbone cyclization. Biosynthetic methods are depicted in green, chemical methods in blue, hybrid strategies are shown in turquoise. References: Intein cyclization: [39–41]; hydrazone cyclization: [43]; proteolytic cyclization: [44,45]; thia-zip cyclization: [46,47].

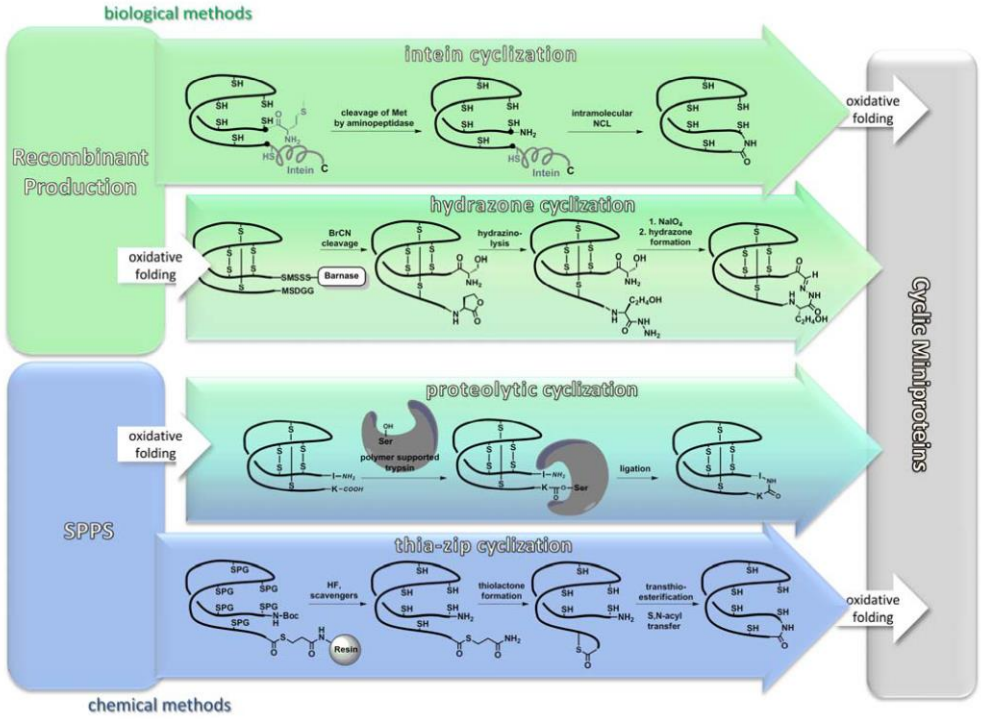


Table 1. Selected folding mixtures for a number of cystine-knot peptides from various families.

Peptide	Type	Folding conditions	Yield/Conversion	Reference
ocMCoTI	ICK	0.5 mM HCl, 200 mM $NaHCO_3$, pH = 9.1, 1–1.5 mg/L peptide.	16% [a] 29% [b]	[48]
cMCoTI	ICK	100 mM NH_4OAc, pH = 8.5, GSH (1-5 mM), 0.1 mg/L peptide	90% [c]	[49]
Variants of ocMCoTI	ICK	50% MeCN in 100 mM $(NH_4)_2CO_3$, GSH (4 eq.)	1.8–7.7% [a] 36–72% [b]	[45]
ICK toxins	ICK	GSSG/GSH (0.3 mM/0.15 mM) in 2 M urea, 100 mM Tris-HCl	3–6% [a]	[50]
EETI-II	ICK	100 mM NH_4OAc, pH = 9.1	>80% [c]	[51]
Gurmarin	ICK	1. Orthogonal cysteine protecting groups 2. GSH/cystamine in 100 mM Tris-HCl, pH = 7.8	1.: 0.55% [a] 2.: 14.1% [a]	[52]
GVIA and analogues	Cono-toxin	Cysteine-selenocysteine exchange, GSSG/GSH (1 mM/2 mM)	60–78% [c]	[53]

Table 1. *Cont.*

Peptide	Type	Folding conditions	Yield/Conversion	Reference
Cycloviolacin O2	CCK	35% DMSO, 6% Brij 35 (an oil dispersant), addition of GSH/cystamine after 24 h (2 mM/ 2 mM) in 100 mM Tris-HCl, pH = 8.5	52% [c]	[54]
Kalata B1	CCK	35% DMSO, 6%, Brij 35 (an oil dispersant), GSH/cystamine (2 mM/ 2 mM) in 100 mM NH_4HCO_3, pH = 8.5	>95% [c]	[31]
Kalata B2	CCK	50% *i*-PrOH, GSSG/GSH (2 mM/2 mM) in 100 mM NH_4HCO_3, pH = 8.5	>95% [c]	[31]
Kalata B8	CCK	50% *i*-PrOH, GSSG/GSH (2 mM/2 mM) in 100 mM NH_4HCO_3, pH = 8.5	>80% [c]	[31]
Cyclic hedyotide B1	CCK	70–80% *i*-PrOH, pH = 8.5 0.033 mg/L peptide	48% [c]	[55]
ASIP	GFCK	100 mM Tris-buffer, pH = 7.7–7.9, 1 mM EDTA, 1 M GuHCl, GSSG/GSH (1:10)	10% [b]	[56]

[a]: yield according to resin loading; [b]: yield according to purified linear precursor; [c]: HPLC conversion.

As ICK peptides do not require any backbone cyclization, they can be recombinantly produced in lower organisms like bacteria or yeast [57–59]. It is important to mention that recombinantly produced cystine-knot peptides can be further chemically modified to yield precursors that contain a non-natural cyclization motif [43]. This issue will be detailed in the following section.

3.2. Chemical Synthesis

SPPS of cysteine-rich peptides has become a routine procedure and peptides comprising more than 30 amino acid residues can be obtained in good yields and enantiomeric purity. From a synthetic point of view, the most challenging issues in SPPS of cyclotides are associated with backbone cyclization (Figure 3). They will be discussed in Section 3.2.3. Nevertheless, chemical synthesis has an obvious advantage over the recombinant route as it allows one to easily incorporate non-natural elements at any desired position in the sequence. Thus, a number of non-canonic building blocks were installed in functional loops of knottins, among them a guaninyl nucleoamino acid as a conformationally restricted and less basic arginine isoster, or homoarginine and amino isobutyric acid that are known to enhance helicity of a peptide chain [48,60]. Furthermore, non-natural elements were inserted in conserved regions of knotted peptides as well. Thus, selenocysteines were installed upon SPPS replacing cysteines at crucial positions of a bracelet cyclotide which resulted in a significant improvement of folding yields [53]. Obviously, installation of non-natural functionalities not only provides an additional option for structural diversity, but also allows for the implementation of coupling sites for backbone cyclization or oligomerization [43,61].

3.2.1. Chain Assembly

SPPS can be conducted by following two different general strategies. In the *tert*-butyloxycarbonyl (Boc) strategy, α-amino groups are protected with acid-labile Boc groups (removed *via* addition of 25% TFA), while deprotection of side chains requires stronger acidic conditions (e.g., HF, methane-sulfonic acid, *etc.*), thus ensuring "pseudo-orthogonality" of the method [62,63]. In the second orthogonal strategy the base-labile fluorenylmethyloxycarbonyl (Fmoc) moiety blocks the α-amino group, whereas side chains can be deprotected with acids (e.g., TFA) [63,64]. To date, Fmoc-SPPS is often the method of choice as less corrosive and aggressive reagents are used and the elongation of the peptide chain during synthesis can be easily monitored at the Fmoc deprotection step [63]. However, the Boc strategy is still applied to SPPS of cystine-knot peptides, as it provides some obvious advantages over the Fmoc strategy [54,55,63]. Besides the incompatibility of Fmoc deprotection with the synthesis of C-terminal thioesters (Section 2.3.2), Boc chemistry often provides higher yields per coupling step [54,55,63]. Furthermore, the prices of Boc-protected amino acids in some cases are still lower in comparison to their Fmoc-protected pendants, although prices for Fmoc-protected amino acids have been decreasing continuously since the introduction of the large-scale industrial synthesis of the HIV fusion inhibitor enfuvirtide (*Fuzeon*®, Roche) [63,65]. The aggregation of growing peptide chains during SPPS dramatically lowers reaction yields [63]. The decreased aggregation tendency of the resin-bound peptide chain that is due to the protonated aminoterminus and backbone resulting from TFA cleavage of aminoterminal Boc groups is one major advantage of Boc- over Fmoc-SPPS [63]. Nevertheless, intermolecular aggregation, the formation of undesired secondary structures and steric hindrance can also be overcome through the usage of microwave irradiation, not only for the raise of the reaction temperature, but also for the polar peptide backbone alignment with the electromagnetic irradiation [63,66]. These effects also lead to increased reaction rates, thereby reducing formation of side-products. Moreover, prolonged reaction times in Fmoc-SPPS are outweighed through the advantage of fully automated synthesizers that can be utilized more regularly due to the usage of less aggressive reagents (although peptide synthesizers compatible with Boc-SPPS are also commercially available). In both methodologies racemization of the amino acid through deprotonating the α-hydrogen with the activator base can be easily overcome by the usage of 2,4,6-tri-methylpyridine or racemization-resistant cysteine protection as e.g., the recently reported 4-methoxy-benzyloxymethyl group [50,67]. In summary, despite Fmoc-SPPS being to date the method of choice, Boc-SPPS is a valuable back-up tool for aggregation-prone peptides or peptides with base-labile moieties which are not compatible with Fmoc-chemistry [54,55,63].

3.2.2. Oxidative Folding

Oxidative folding of linear or head-to-tail cyclized precursors towards the bioactive isomer is the most important and also most critical step during synthesis of cystine-knot peptides. Significant efforts have been made to determine folding pathways and optimize oxidative folding conditions [30,31,36,49,68–73].

Particularly the optimization of folding conditions of cyclotides is challenging since they contain patches of hydrophobic residues on their surface, which substantially contribute to their

membrane-disrupting activity [11]. These highly aggregation-prone residues tend to stick together in a non-native conformation, making therefore preorganization *via* backbone cyclization essential to obtain acceptable yields in the folding process [16,74]. CCK peptides belonging to the Möbius or bracelet family, respectively, follow different folding pathways. Möbius cystine-knot peptides rapidly form and accumulate an energetically trapped two-cystine intermediate which lacks the penetrating III,VI-disulfide, and finally fold in the native conformation either directly or *via* another, non-native three-cystine intermediate (Figure 4) [31,36].

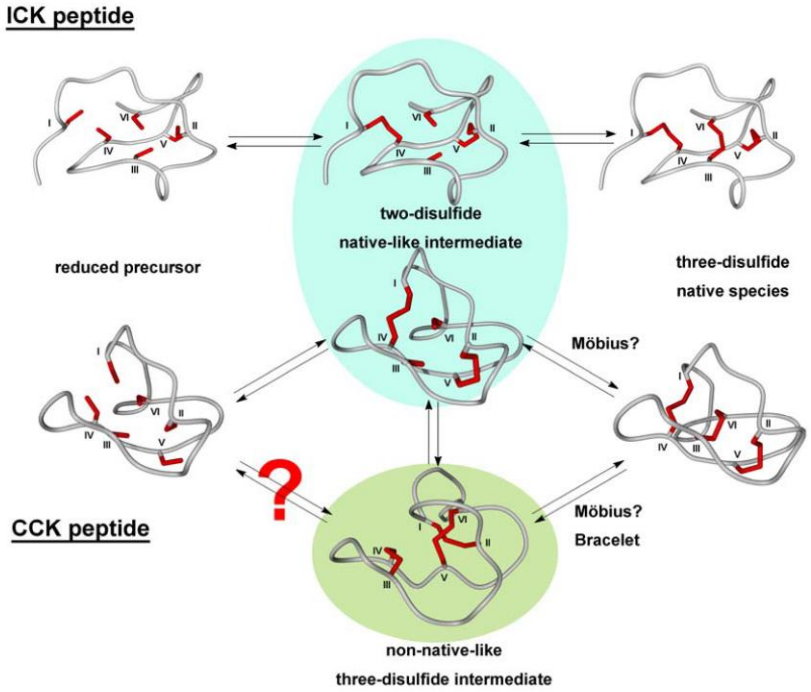
The folding pathway varies with the addition of redox assistants or organic solvents [31]. However, in contrast to the members of the bracelet family, Möbius CCK peptides fold into the bioactive conformation spontaneously in good to moderate yields [31,36,38]. The *in vitro* folding of bracelet cyclotides is more challenging as their kinetic trap is not a two-cystine, but the non-native CysI-CysII, CysIII-CysIV, CysV-CysVI "ladder-like" isomer (Figure 4) [31,54]. Thus, the addition of accessory redox agents is essential. However, no recipe has been reported to date, allowing for the formation of the native form as major product [31]. To overcome these problems, cyclotides, particularly those that cannot be obtained in reasonable yields *via* random cysteine oxidation, were subjected to a stepwise folding procedure with consecutive cleavage of orthogonally protected cysteine side chains or *via* the incorporation of selenocysteines as diselenides possess a higher reduction potential [53,74,75]. Orthogonal oxidation of multiple cysteine pairs results in selective disulfide formation, thus ensuring the desired cystine pattern. From the huge repertoire of cysteine side-chain protecting groups, trityl, acetamidomethyl, S-*tert*-butyl, and *tert*-butyl are the most often used orthogonal combinations [76–78]. A number of cystine-knot peptides were synthesized following the strategy of regioselective disulfide formation, among them anti-malaria peptides from *Psalmopoeus cambridgei* and the ICK peptide gurmarin [52,79].

Removal of a non-critical disulfide bridge and substitution of remaining disulfides with diselenides is another way to simplify folding of multidisulfide peptides and was successfully applied to the synthesis of some conotoxines [67,80]. Interestingly, for the peptides possessing a two-disulfide pattern both cystines could be replaced on-support with isosteric cystathionine motifs without loss of bioactivity [81].

Common to twisted cyclotides, the folding pathway of ICK peptides follows a direct route resulting in a quickly formed, kinetically trapped two-disulfide intermediate [36,49,82]. This intermediate consists of two "low-energy" disulfides and therefore is vulnerable for misfolding. Thus, peptide chain preorganization and formation of correctly folded intermediates remarkably contribute to the yield and purity of the final bioactive isomer. In contrast to Möbius cyclotides, the formation of the third ring-penetrating cystine directly yields the native conformation (Figure 4) [49].

The increased hydrophilicity of the active loops in ICK peptides admits a higher degree of structural preorganization compared to the mainly hydrophobic CCK peptides, thus enhancing their folding yields. Hence, ICK peptides frequently are not macrocyclic. Nevertheless, for MCoTI variants the cyclic form displayed an improved folding behavior compared to the linear variant [27,49].

Figure 4. *In vitro* folding pathways of cystine-knot peptides from different families. Most ICK peptides are thought to proceed to the folded form *via* formation of two-disulfide native-like intermediate. Folding of CCK peptides may either follow a similar path or proceed *via* a non-native 3-cystine intermediate. Structures were modeled and energy-minimized with Yasara Ver. 12.4.1.



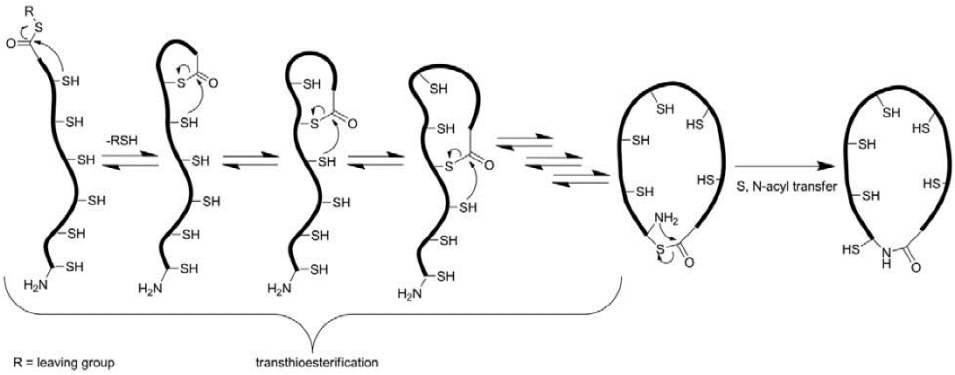
Many different folding assistants have been used in various combinations, each optimized for an individual protein or peptide. Common requirements for efficient cystine-knot peptide folding are high dilution, significant ratios of organic solvents (e.g., DMSO, *i*-PrOH) and presence of redox folding assistants (e.g., ox./red. glutathione) [30,31,36,49,68–72]. Despite the importance of the oxidative folding of cystine-knot peptides, particularly of those that contain grafted sequences, yields are rarely given in the present literature [54,57–59,83]. Moreover, missing distinctions between yield and HPLC-observed conversion rates and indications, whether they are based on resin loading or correspond to the crude or purified precursor further complicate the summarized comparison of folding yields (Table 1).

3.2.3. Backbone Macrocyclization

Backbone cyclization usually is accomplished *via* a so-called 'thia-zip' mechanism using the native chemical ligation (NCL) technology (Figure 5) [46,47,69,84–87]. To this end, a carboxyterminal leaving group, generally a thioester, must be installed. Introduction of this moiety can be achieved through thioesterification of the fully protected peptide in solution, either as a cleavable linker on the

peptide resin or as a reagent during nucleophilic cleavage. For the incorporation of the thioester after chain assembly, the peptide has to be synthesized on an 'ultra-acid-labile' resin (e.g., a TGT resin) from which the peptide can be cleaved with all side-chain protecting groups intact. To that fully protected peptide the thiol is coupled forming the desired thioester [88]. Although this methodology is compatible with the common Fmoc-strategy of peptide synthesis, it is subjected to imponderabilities due to the unpredictable solubility of fully protected peptides, especially peptides of that size. Moreover, undesired carboxyterminal racemization may occur during synthesis [88,89]. This problem can be overcome very elegantly by choosing a glycine as aminoterminal and a cysteine as carboxyterminal residue as the site of macrocyclization, because glycine is the only non-chiral amino acid and glycine-cysteine combinations exist in a number of cystine-knot peptides (Figure 1) [88]. Installation of a thioester as on-resin cleavable linker seems more elegant, as no special modification is needed [86]. Unfortunately, piperidine that is a common reagent used in Fmoc-SPPS for N-terminal deprotection is not compatible with that linker as its nucleophilic attack at the thioester results in the cleavage of the peptide chain from the resin [86]. A combination of non-nucleophilic 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) and 1-hydroxybenzotriazole (HOBt) helps to overcome that problem on the cost of an enhanced aspartimide formation [90]. Interestingly, it has not been checked so far, whether 2-methylpiperidine could solve these problems as its utility for the synthesis of peptides with piperidine-labile tyrosine sulfate esters was demonstrated [91]. The third possibility, though not yet elaborated, might be the usage of safety-catch linkers (e.g., hydrazinobenzoyl) which can be cleaved by a respective nucleophile after suitable activation [92,93].

Figure 5. Proposed cyclization *via* native chemical ligation.



After successful incorporation of a carboxyterminal thioester, NCL frequently provides cyclic peptides in excellent conversions or yields, respectively (Table 2) [86,88]. Although to date the mechanism is not fully understood and not all intermediates are precisely characterized, it is commonly accepted that intramolecular thioesterifications of the internal thiol groups and the carboxyterminus take place (Figure 5) [46]. This "thia-zip" rearrangement gradually increases ring size and eventually brings both termini in close proximity. As a consequence, an irreversible *S*, *N*-acyl transfer is induced, finally leading to the cyclic product [46,84]. This model is supported by various studies, in which the aminoterminus was acetylated, a linker introduced or the ring-chain tautomeric equilibrium

investigated [46,84,94]. Nevertheless, NCL-driven macrocyclization has been reported also for peptides possessing only an aminoterminal cysteine. Due to the lack of multiple thiol groups, intramolecular thialactone exchange is not possible in such molecules and ring closure takes place without zip-like rearrangements [84–86,95].

Despite the success of NCL as the method of choice, two alternative backbone cyclization strategies have been reported (Table 2) [43,44]. One method relies on the bacterial production of the cystine-knot peptide in *E. coli via* fusion to a carrier protein [43]. Therein, the linear precursor is fused to barnase, an RNAse from *Bacillus amyloliquefaciens*, which guides the fused protein complex into the periplasm of the Gram-negative *E. coli*, where the oxidative milieu supports folding [43,96]. Chemical head-to-tail cyclization of the folded ICK peptide was achieved through the formation of a stable *N*-C hydrazone linkage between a periodate-oxidized aminoterminal serine and a carboxyterminal hydrazide. This moiety was generated by hydrazinolysis of a homoserine lactone formed upon cyanogen bromide cleavage at a unique methionine that was present at the junction of the knottin and the carrier protein sequence (Figure 3) [43]. Recently, protease-mediated backbone cyclization was accomplished using immobilized trypsin [44,45]. Therein, a solution of chemically synthesized and correctly folded MCoTI-II was added to polymer-bound trypsin and, upon covalent binding to the active site of this protease, C- and N-termini of the cystine knot were brought into close proximity and ligated between the P1 lysine and P1' leucine within the protease inhibiting loop (Figure 3) [44,45,97]. Conversion rates and yields for the various cyclization methods are summarized in Table 2.

Table 2. Selected cyclization conditions for a number of cystine-knot peptides from various families.

Peptide	Type	Cyclization reaction; conditions	Yield/Conversion	Reference
Variants of cMCoTI	ICK	Immobilized trypsin; 100 mM phosphate buffer, pH = 7.4	90–94% [a]	[44,45]
Variants of cMCoTI	ICK	Folding and NCL as one-pot reaction; 50% GSSG (1 mM) in 100 mM carbonate buffer, 50% peptide (3 mM) in acetonitrile	63–72% [b]	[45]
Variants of cMCoTI	ICK	Hydrazone linkage; multiple reactions from recombinantly produced barnase fusion	0.5–1 mg/L [c]	[43]
Kalata B1	CCK	Amide bond; HBTU or BOP, respectively (1–10 eq.), DIEA in DMF	~25% [a]	[69]
Kalata B1	CCK	NCL; 100 mM NaH_2PO_4, TCEP (6 eq.), pH = 7.4, 1 mg/mL peptide	100% [a]	[69]
hB1	CCK	NCL; 100 mM NaH_2PO_4, 6 M GuHCl, thiophenol (100 eq.), pH = 7.5	100% [a]	[55]
Cyclic MrIA	Cyclic cono-toxin	NCL; 100 mM Tris-HCl, pH = 7.8, 6 M GuHCl, sodium 2-sulfonylethane sulfonate (1 mg/mL), anaerobic	100% [a]	[98]

[a]: HPLC conversion; [b]: yield according to purified precursor; [c]: yield of purified cystine-knot peptide is given per liter cell culture.

3.2.4. Analysis of Cystine Knots

RP-HPLC in combination with mass spectrometry, especially ESI-MS and MALDI-TOF, are commonly used for the routine analysis of cystine-knot peptides [23,31,48,50,55]. Therein, not only the polarity, but also the molecular weight are determined giving clear evidence of the quality and nature of the product [23,31,48,50,55]. For example, the progress of oxidative folding was determined through a shift in RP-HPLC retention time as well as a decreased molecular weight because of the loss of the respective number of hydrogens [23,31,48,55]. Unfortunately, this is not the final proof of correct folding as topology of the cystine connections is essential for bioactivity [51]. Therefore, determination of the correct disulfide topology is necessary which is applied routinely *via* MS-MS analytics, Edman sequencing, and protein digestion followed by subsequent MS analysis of the reaction mixture [21,24,55,95,99]. From the resulting fragments cystine connections can be deduced. Due to the small size of cystine-knot peptides, detailed structural information as the connectivity of hydrogen bonds has been collected by 2D NMR studies [25,27,82]. Nevertheless, although to date only the crystal structure of a cyclotide and an open-chain knottin have been solved, ICK peptides can be easily co-crystallized with their target enzymes [2,3,12,100].

4. Applications to Drug Design

Potential applications of cystine-knot peptides to drug design have been extensively reviewed [1,5,10,101]. In this section a few examples of recently developed peptides are given that highlight the advantages of their use for *in vivo* and *in vitro* targeting of disease-related molecules.

Cystine-knot peptides possess three essential characteristics which are desirable for the application as drugs: excellent stability, high affinity or inhibitory activity, and the potential for high selectivity towards the target. Additionally, these three-disulfide scaffolds provide a remarkable sequence tolerance allowing for the introduction of novel functionalities within their loop region, often without the loss of structural integrity and bioactivity [10]. Moreover, at least members of the ICK family are thought to have no cytotoxic properties and demonstrate good body clearance and tissue distribution, although these characteristics require validation for each modified candidate [1,10,101]. Interestingly, some cystine-knot peptides have been reported to be cell-penetrating [97,102].

The choice of ICK or CCK peptides as a scaffold for drug design is mainly guided by the natural target molecule, although several instances of engineering towards completely different targets have been reported [59,103,104]. As natural ICK peptides usually target trypsin-like proteases, they are optimal starting points for obtaining potent inhibitors against other disease-related serine proteases [9,23,26]. Engineered "imino-cyclotides" combining the ICK backbone with a non-natural hydrazone cyclization motif have been reported to inhibit human mast cell tryptase β, a protease of interest as a therapeutic target for the treatment of inflammatory disorders and allergic asthma [43].

For tumor targeting, several transmembrane serine proteases which are known to be overexpressed in cancer cells are also valuable targets [105–108].

Most interestingly, Agouti-related miniprotein (AgRP), an acyclic four-cystine knottin, has recently been modified with an RGD peptide motif towards binding of cancer-dependent integrins and the resulting constructs were used for radio imaging *in vivo* [59,103,109]. A prototype for the engineering

of miniproteins towards variants with antiviral activity is the HIV entry inhibitor CD4M47[Phe]. Here, the miniprotein Leiurotoxin I from the deathstalker scorpion *Leiurus quinquestriatus hebraeus* was used as a structural scaffold [110–113]. Several rounds of directed evolution and rational design resulted in an optimized binding towards gp120 of the viral particle of HIV, thus inhibiting cell entry [110–112]. CCK peptides in most cases have shown antiviral or bactericidal activity in their wild-type form, but to date only some of them have been grafted towards new bioactivities [6,20,114,115].

5. Conclusions and Outlook

Cystine-knot peptides are defined through their unique architecture which endows them with an extremely high stability and sequence tolerance resulting in promising scaffolds for drug development and chemical genetics. Current synthetic problems, oxidative folding and backbone cyclization, depend on whether cystine-knot peptides are recombinantly produced or chemically synthesized. Head-to-tail macrocyclization is problematic for recombinantly produced peptides, and formation of the three-disulfide pattern for those chemically synthesized. As for large-scale industrial processes, *in vivo* synthesis may become a cost-effective alternative to chemical synthesis, but microorganisms and their respective production conditions need further optimization. Novel pharmaceuticals based on cystine-knot peptides may find their way to clinical trials in the next couple of years. Continuously reported improvements in their functionalization towards modulators of disease-relevant targets in combination with the increasing number of publications for both chemical synthesis and recombinant production provide excellent future prospects.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft through grant KO 1390/10-1 “Chemoselective reactions for the synthesis and application of functional proteins”.

References

1. Kolmar, H. Natural and engineered cystine knot miniproteins for diagnostic and therapeutic applications. *Curr. Pharm. Des.* **2011**, *17*, 4329–4336.
2. Gelly, J.C.; Gracy, J.; Kaas, Q.; Le-Nguyen, D.; Heitz, A.; Chiche, L. The KNOTTIN website and database: A new information system dedicated to the knottin scaffold. *Nucleic Acids Res.* **2004**, *32*, D156–D159.
3. Gracy, J.; Le-Nguyen, D.; Gelly, J.C.; Kaas, Q.; Heitz, A.; Chiche, L. KNOTTIN: The knottin or inhibitor cystine knot scaffold in 2007. *Nucleic Acids Res.* **2008**, *36*, D314–D319.
4. Craik, D.J.; Daly, N.L.; Waine, C. The cystine knot motif in toxins and implications for drug design. *Toxicon* **2001**, *39*, 43–60.
5. Daly, N.L.; Craik, D.J. Bioactive cystine knot proteins. *Curr. Opin. Chem. Biol.* **2011**, *15*, 362–368.
6. Gran, L. On the effect of a polypeptide isolated from “Kalata-Kalata” (Oldenlandia affinis DC) on the oestrogen dominated uterus. *Acta Pharmacol. Toxicol. (Copenh.)* **1973**, *33*, 400–408.

7. McDonald, N.Q.; Lapatto, R.; Murray-Rust, J.; Gunning, J.; Wlodawer, A.; Blundell, T.L. New protein fold revealed by a 2.3-A resolution crystal structure of nerve growth factor. *Nature* **1991**, *354*, 411–414.
8. Felizmenio-Quimio, M.E.; Daly, N.L.; Craik, D.J. Circular proteins in plants: Solution structure of a novel macrocyclic trypsin inhibitor from Momordica cochinchinensis. *J. Biol. Chem.* **2001**, *276*, 22875–22882.
9. Chiche, L.; Heitz, A.; Gelly, J.C.; Gracy, J.; Chau, P.T.; Ha, P.T.; Hernandez, J.F.; Le-Nguyen, D. Squash inhibitors: From structural motifs to macrocyclic knottins. *Curr. Protein Pept. Sci.* **2004**, *5*, 341–349.
10. Kolmar, H. Biological diversity and therapeutic potential of natural and engineered cystine knot miniproteins. *Curr. Opin. Pharmacol.* **2009**, *9*, 608–614.
11. Craik, D.J.; Cemazar, M.; Wang, C.K.; Daly, N.L. The cyclotide family of circular miniproteins: Nature's combinatorial peptide template. *Biopolymers* **2006**, *84*, 250–266.
12. Wang, C.K.; Hu, S.H.; Martin, J.L.; Sjogren, T.; Hajdu, J.; Bohlin, L.; Claeson, P.; Goransson, U.; Rosengren, K.J.; Tang, J.; *et al.* Combined X-ray and NMR analysis of the stability of the cyclotide cystine knot fold that underpins its insecticidal activity and potential use as a drug scaffold. *J. Biol. Chem.* **2009**, *284*, 10672–10683.
13. Craik, D.J.; Daly, N.L. NMR as a tool for elucidating the structures of circular and knotted proteins. *Mol. Biosyst.* **2007**, *3*, 257–265.
14. Colgrave, M.L.; Craik, D.J. Thermal, Chemical, And enzymatic stability of the cyclotide kalata B1: The importance of the cyclic cystine knot. *Biochemistry* **2004**, *43*, 5965–5975.
15. Jennings, C.; West, J.; Waine, C.; Craik, D.; Anderson, M. Biosynthesis and insecticidal properties of plant cyclotides: The cyclic knotted proteins from Oldenlandia affinis. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10614–10619.
16. Tam, J.P.; Lu, Y.A.; Yang, J.L.; Chiu, K.W. An unusual structural motif of antimicrobial peptides containing end-to-end macrocycle and cystine-knot disulfides. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 8913–8918.
17. Gustafson, K.R.; Sowder, R.C.; Henderson, L.E.; Parsons, I.C.; Kashman, Y.; Cardellina, J.H.; McMahon, J.B.; Buckheit, R.W.; Pannell, L.K.; Boyd, M.R. Circulins A and B, Novel human immunodeficiency virus (HIV)-inhibitory macrocyclic peptides from the tropical tree Chassalia parvifolia. *J. Am. Chem. Soc.* **1994**, *116*, 9337–9338.
18. Svangard, E.; Goransson, U.; Hocaoglu, Z.; Gullbo, J.; Larsson, R.; Claeson, P.; Bohlin, L. Cytotoxic cyclotides from Viola tricolor. *J. Nat. Prod.* **2004**, *67*, 144–147.
19. Lindholm, P.; Goransson, U.; Johansson, S.; Claeson, P.; Gullbo, J.; Larsson, R.; Bohlin, L.; Backlund, A. Cyclotides: A novel type of cytotoxic agents. *Mol. Cancer Ther.* **2002**, *1*, 365–369.
20. Gustafson, K.R.; McKee, T.C.; Bokesch, H.R. Anti-HIV cyclotides. *Curr. Protein Pept. Sci.* **2004**, *5*, 331–340.
21. Hernandez, J.F.; Gagnon, J.; Chiche, L.; Nguyen, T.M.; Andrieu, J.P.; Heitz, A.; Trinh Hong, T.; Pham, T.T.; Le Nguyen, D. Squash trypsin inhibitors from Momordica cochinchinensis exhibit an atypical macrocyclic structure. *Biochemistry* **2000**, *39*, 5722–5730.

22. Heitz, A.; Hernandez, J.F.; Gagnon, J.; Hong, T.T.; Pham, T.T.; Nguyen, T.M.; Le-Nguyen, D.; Chiche, L. Solution structure of the squash trypsin inhibitor MCoTI-II: A new family for cyclic knottins. *Biochemistry* **2001**, *40*, 7973–7983.
23. Avrutina, O.; Schmoldt, H.U.; Gabrijelcic-Geiger, D.; Le Nguyen, D.; Sommerhoff, C.P.; Diederichsen, U.; Kolmar, H. Trypsin inhibition by macrocyclic and open-chain variants of the squash inhibitor MCoTI-II. *Biol. Chem.* **2005**, *386*, 1301–1306.
24. Kowalska, J.; Pszczola, K.; Wilimowska-Pelc, A.; Lorenc-Kubis, I.; Zuziak, E.; Lugowski, M.; Legowska, A.; Kwiatkowska, A.; Sleszynska, M.; Lesner, A.; *et al*. Trypsin inhibitors from the garden four o'clock (Mirabilis jalapa) and spinach (*Spinacia oleracea*) seeds: Isolation, Characterization and chemical synthesis. *Phytochemistry* **2007**, *68*, 1487–1496.
25. Heitz, A.; Chiche, L.; Le-Nguyen, D.; Castro, B. 1H 2D NMR and distance geometry study of the folding of Ecballium elaterium trypsin inhibitor, A member of the squash inhibitors family. *Biochemistry* **1989**, *28*, 2392–2398.
26. Favel, A.; Mattras, H.; Coletti-Previero, M.A.; Zwilling, R.; Robinson, E.A.; Castro, B. Protease inhibitors from Ecballium elaterium seeds. *Int. J. Pept. Protein Res.* **1989**, *33*, 202–208.
27. Heitz, A.; Avrutina, O.; Le-Nguyen, D.; Diederichsen, U.; Hernandez, J.F.; Gracy, J.; Kolmar, H.; Chiche, L. Knottin cyclization: Impact on structure and dynamics. *BMC Struct. Biol.* **2008**, *8*, 54.
28. Clark, R.J.; Craik, D.J. Native chemical ligation applied to the synthesis and bioengineering of circular peptides and proteins. *Biopolymers* **2010**, *94*, 414–422.
29. Conlan, B.F.; Anderson, M.A. Circular micro-proteins and mechanisms of cyclization. *Curr. Pharm. Des.* **2011**, *17*, 4318–4328.
30. Craik, D.J. The folding of disulfide-rich proteins. *Antioxid. Redox. Signal.* **2011**, *14*, 61–64.
31. Aboye, T.L.; Clark, R.J.; Burman, R.; Roig, M.B.; Craik, D.J.; Goransson, U. Interlocking disulfides in circular proteins: Toward efficient oxidative folding of cyclotides. *Antioxid. Redox. Signal.* **2011**, *14*, 77–86.
32. Schieck, A.; Muller, T.; Schulze, A.; Haberkorn, U.; Urban, S.; Mier, W. Solid-phase synthesis of the lipopeptide Myr-HBVpreS/2–78, A hepatitis B virus entry inhibitor. *Molecules* **2010**, *15*, 4773–4783.
33. Boulègue, C.; Musiol, H.J.; Prasas, V.; Moroder, L. Synthesis of cystine-rich peptides. *Chem. Today* **2006**, *24*, 24–36.
34. Zhu, S.; Darbon, H.; Dyason, K.; Verdonck, F.; Tytgat, J. Evolutionary origin of inhibitor cystine knot peptides. *FASEB J.* **2003**, *17*, 1765–1767.
35. Goransson, U.; Craik, D.J. Disulfide mapping of the cyclotide kalata B1: Chemical proof of the cystic cystine knot motif. *J. Biol. Chem.* **2003**, *278*, 48188–48196.
36. Daly, N.L.; Clark, R.J.; Craik, D.J. Disulfide folding pathways of cystine knot proteins: Tying the knot within the circular backbone of the cyclotides. *J. Biol. Chem.* **2003**, *278*, 6314–6322.
37. Ireland, D.C.; Wang, C.K.; Wilson, J.A.; Gustafson, K.R.; Craik, D.J. Cyclotides as natural anti-HIV agents. *Biopolymers* **2008**, *90*, 51–60.
38. Craik, D.J.; Conibear, A.C. The chemistry of cyclotides. *J. Org. Chem.* **2011**, *76*, 4805–4817.
39. Austin, J.; Wang, W.; Puttamadappa, S.; Shekhtman, A.; Camarero, J.A. Biosynthesis and biological screening of a genetically encoded library based on the cyclotide MCoTI-I. *ChemBioChem* **2009**, *10*, 2663–2670.

40. Camarero, J.A.; Kimura, R.H.; Woo, Y.H.; Shekhtman, A.; Cantor, J. Biosynthesis of a fully functional cyclotide inside living bacterial cells. *ChemBioChem* **2007**, *8*, 1363–1366.
41. Kimura, R.H.; Tran, A.T.; Camarero, J.A. Biosynthesis of the cyclotide Kalata B1 by using protein splicing. *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 973–976.
42. Puttamadappa, S.S.; Jagadish, K.; Shekhtman, A.; Camarero, J.A. Backbone dynamics of cyclotide MCoTI-I free and complexed with trypsin. *Angew. Chem. Int. Ed. Engl.* **2010**, *49*, 7030–7034.
43. Avrutina, O.; Schmoldt, H.U.; Gabrijelcic-Geiger, D.; Wentzel, A.; Frauendorf, H.; Sommerhoff, C.P.; Diederichsen, U.; Kolmar, H. Head-to-tail cyclized cystine-knot peptides by a combined recombinant and chemical route of synthesis. *ChemBioChem* **2008**, *9*, 33–37.
44. Thongyoo, P.; Jaulent, A.M.; Tate, E.W.; Leatherbarrow, R.J. Immobilized protease-assisted synthesis of engineered cysteine-knot microproteins. *ChemBioChem* **2007**, *8*, 1107–1109.
45. Thongyoo, P.; Roque-Rosell, N.; Leatherbarrow, R.J.; Tate, E.W. Chemical and biomimetic total syntheses of natural and engineered MCoTI cyclotides. *Org. Biomol. Chem.* **2008**, *6*, 1462–1470.
46. Dawson, P.E.; Muir, T.W.; Clark-Lewis, I.; Kent, S.B. Synthesis of proteins by native chemical ligation. *Science* **1994**, *266*, 776–779.
47. Lu, W.; Starovasnik, M.A.; Kent, S.B. Total chemical synthesis of bovine pancreatic trypsin inhibitor by native chemical ligation. *FEBS Lett.* **1998**, *429*, 31–35.
48. Avrutina, O.; Schmoldt, H.-U.; Kolmar, H.; Diederichsen, U. Fmoc-Assisted Synthesis of a 29-Residue Cystine-Knot Trypsin Inhibitor Containing a Guaninyl Amino Acid at the P1-Position. *Eur. J. Org. Chem.* **2004**, *2004*, 4931–4935.
49. Cemazar, M.; Daly, N.L.; Haggblad, S.; Lo, K.P.; Yulyaningsih, E.; Craik, D.J. Knots in rings, The circular knotted protein Momordica cochinchinensis trypsin inhibitor-II folds via a stable two-disulfide intermediate. *J. Biol. Chem.* **2006**, *281*, 8224–8232.
50. Park, J.H.; Carlin, K.P.; Wu, G.; Ilyin, V.I.; Kyle, D.J. Cysteine racemization during the Fmoc solid phase peptide synthesis of the Nav1.7-selective peptide—protoxin II. *J. Pept. Sci.* **2012**, *18*, 442–448.
51. Wentzel, A.; Christmann, A.; Kratzner, R.; Kolmar, H. Sequence requirements of the GPNG beta-turn of the Ecballium elaterium trypsin inhibitor II explored by combinatorial library screening. *J. Biol. Chem.* **1999**, *274*, 21037–21043.
52. Eliasen, R.; Andresen, T.L.; Conde-Frieboes, K.W. Handling a tricycle: Orthogonal versus random oxidation of the tricyclic inhibitor cystine knotted peptide gurmarin. *Peptides* **2012**, *37*, 144–149.
53. Gowd, K.H.; Yarotskyy, V.; Elmslie, K.S.; Skalicky, J.J.; Olivera, B.M.; Bulaj, G. Site-specific effects of diselenide bridges on the oxidative folding of a cystine knot peptide, omega-selenoconotoxin GVIA. *Biochemistry* **2010**, *49*, 2741–2752.
54. Leta Aboye, T.; Clark, R.J.; Craik, D.J.; Goransson, U. Ultra-stable peptide scaffolds for protein engineering-synthesis and folding of the circular cystine knotted cyclotide cycloviolacin O2. *ChemBioChem* **2008**, *9*, 103–113.
55. Wong, C.T.; Taichi, M.; Nishio, H.; Nishiuchi, Y.; Tam, J.P. Optimal oxidative folding of the novel antimicrobial cyclotide from Hedyotis biflora requires high alcohol concentrations. *Biochemistry* **2011**, *50*, 7275–7283.

56. McNulty, J.C.; Jackson, P.J.; Thompson, D.A.; Chai, B.; Gantz, I.; Barsh, G.S.; Dawson, P.E.; Millhauser, G.L. Structures of the agouti signaling protein. *J. Mol. Biol.* **2005**, *346*, 1059–1070.
57. Sommerhoff, C.P.; Avrutina, O.; Schmoldt, H.U.; Gabrijelcic-Geiger, D.; Diederichsen, U.; Kolmar, H. Engineered cystine knot miniproteins as potent inhibitors of human mast cell tryptase beta. *J. Mol. Biol.* **2010**, *395*, 167–175.
58. Kimura, R.H.; Jones, D.S.; Jiang, L.; Miao, Z.; Cheng, Z.; Cochran, J.R. Functional mutation of multiple solvent-exposed loops in the Ecballium elaterium trypsin inhibitor-II cystine knot miniprotein. *PLoS One* **2011**, *6*, e16112.
59. Silverman, A.P.; Kariolis, M.S.; Cochran, J.R. Cystine-knot peptides engineered with specificities for alpha(IIb)beta(3) or alpha(IIb)beta(3) and alpha(v)beta(3) integrins are potent inhibitors of platelet aggregation. *J. Mol. Recognit.* **2011**, *24*, 127–135.
60. Blind, M.; Kolmar, H. Polypeptides comprising a knottin protein moiety. Patent EP 1,958,957,A1, 18 February 2008.
61. Fabritz, S.; Horner, S.; Konning, D.; Empting, M.; Reinwarth, M.; Dietz, C.; Glotzbach, B.; Frauendorf, H.; Kolmar, H.; Avrutina, O. From pico to nano: Biofunctionalization of cube-octameric silsesquioxanes by peptides and miniproteins. *Org. Biomol. Chem.* **2012**, *10*, 6287–6293.
62. Merrifield, R.B. Solid-Phase Peptide Synthesis. 3. An Improved Synthesis of Bradykinin. *Biochemistry* **1964**, *3*, 1385–1390.
63. Kates, S.A.; Albericio, F. *Solid-Phase Synthesis: A Practical Guide*, 1st ed.; CRC Press: Boca Raton, FL, USA, 2000; Volume 1, p. 848.
64. Chang, C.D.; Meienhofer, J. Solid-phase peptide synthesis using mild base cleavage of N alpha-fluorenylmethyloxycarbonylamino acids, exemplified by a synthesis of dihydrosomatostatin. *Int. J. Pept. Protein Res.* **1978**, *11*, 246–249.
65. Robertson, D. US FDA approves new class of HIV therapeutics. *Nat. Biotechnol.* **2003**, *21*, 470–471.
66. Kappe, C.O.; Stadler, A. *Microwaves in Organic and Medicinal Chemistry*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2005; p. 409.
67. Hibino, H.; Nishiuchi, Y. 4-Methoxybenzyloxymethyl group, a racemization-resistant protecting group for cysteine in Fmoc solid phase peptide synthesis. *Org. Lett.* **2012**, *14*, 1926–1929.
68. Cemazar, M.; Gruber, C.W.; Craik, D.J. Oxidative folding of cyclic cystine knot proteins. *Antioxid. Redox. Signal.* **2008**, *10*, 103–111.
69. Daly, N.L.; Love, S.; Alewood, P.F.; Craik, D.J. Chemical synthesis and folding pathways of large cyclic polypeptides: Studies of the cystine knot polypeptide kalata B1. *Biochemistry* **1999**, *38*, 10606–10614.
70. Gunasekera, S.; Daly, N.L.; Clark, R.J.; Craik, D.J. Dissecting the oxidative folding of circular cystine knot miniproteins. *Antioxid. Redox. Signal.* **2009**, *11*, 971–980.
71. Muller, C.; Richter, S.; Rinas, U. Kinetics control preferential heterodimer formation of platelet-derived growth factor from unfolded A- and B-chains. *J. Biol. Chem.* **2003**, *278*, 18330–18335.
72. Cemazar, M.; Joshi, A.; Daly, N.L.; Mark, A.E.; Craik, D.J. The structure of a two-disulfide intermediate assists in elucidating the oxidative folding pathway of a cyclic cystine knot protein. *Structure* **2008**, *16*, 842–851.
73. Green, B.R.; Bulaj, G. Oxidative folding of conotoxins in immobilized systems. *Protein Pept. Lett.* **2006**, *13*, 67–70.

74. Tam, J.P.; Lu, Y.A. A biomimetic strategy in the synthesis and fragmentation of cyclic protein. *Protein Sci.* **1998**, *7*, 1583–1592.
75. Steiner, A.M.; Bulaj, G. Optimization of oxidative folding methods for cysteine-rich peptides: A study of conotoxins containing three disulfide bridges. *J. Pept. Sci.* **2011**, *17*, 1–7.
76. Isidro-Llobet, A.; Álvarez, M.; Albericio, F. Amino Acid-Protecting Groups. *Chem. Rev.* **2009**, *109*, 2455–2504.
77. Veber, D.F.; Milkowski, J.D.; Varga, S.L.; Denkewalter, R.G.; Hirschmann, R. Acetamidomethyl, A novel thiol protecting group for cysteine. *J. Am. Chem. Soc.* **1972**, *94*, 5456–5461.
78. Zervas, L.; Photaki, I. On Cysteine and Cystine Peptides. I. New S-Protecting Groups for Cysteine. *J. Am. Chem. Soc.* **1962**, *84*, 3887–3897.
79. Kamolkijkarn, P.; Prasertdee, T.; Netirojjanakul, C.; Sarnpitak, P.; Ruchirawat, S.; Deechongkit, S. Synthesis, Biophysical, And biological studies of wild-type and mutant psalmopeotoxins—anti-malarial cysteine knot peptides from Psalmopoeus cambridgei. *Peptides* **2010**, *31*, 533–540.
80. Raffa, R.B. Diselenium, instead of disulfide, bonded analogs of conotoxins: Novel synthesis and pharmacotherapeutic potential. *Life Sci.* **2010**, *87*, 451–456.
81. Dekan, Z.; Vetter, I.; Daly, N.L.; Craik, D.J.; Lewis, R.J.; Alewood, P.F. alpha-Conotoxin ImI incorporating stable cystathionine bridges maintains full potency and identical three-dimensional structure. *J. Am. Chem. Soc.* **2011**, *133*, 15866–15869.
82. Le-Nguyen, D.; Heitz, A.; Chiche, L.; el Hajji, M.; Castro, B. Characterization and 2D NMR study of the stable [9–21, 15–27] 2 disulfide intermediate in the folding of the 3 disulfide trypsin inhibitor EETI II. *Protein Sci.* **1993**, *2*, 165–174.
83. Krause, S.; Schmoldt, H.U.; Wentzel, A.; Ballmaier, M.; Friedrich, K.; Kolmar, H. Grafting of thrombopoietin-mimetic peptides into cystine knot miniproteins yields high-affinity thrombopoietin antagonists and agonists. *FEBS J.* **2007**, *274*, 86–95.
84. Zhang, L.; Tam, J.P. Synthesis and Application of Unprotected Cyclic Peptides as Building Blocks for Peptide Dendrimers. *J. Am. Chem. Soc.* **1997**, *119*, 2363–2370.
85. Camarero, J.A.; Muir, T.W. Chemoselective backbone cyclization of unprotected peptides. *Chem. Commun.* **1997**, 1369–1370.
86. Camarero, J.A.; Cotton, G.J.; Adeva, A.; Muir, T.W. Chemical ligation of unprotected peptides directly from a solid support. *J. Pept. Res.* **1998**, *51*, 303–316.
87. Tam, J.P.; Lu, Y.-A.; Yu, Q. Thia Zip Reaction for Synthesis of Large Cyclic Peptides: Mechanisms and Applications. *J. Am. Chem. Soc.* **1999**, *121*, 4316–4324.
88. Park, S.; Gunasekera, S.; Aboye, T.; Göransson, U. An Efficient Approach for the Total Synthesis of Cyclotides by Microwave Assisted Fmoc-SPPS. *Int. J. Pept. Res. Ther.* **2010**, *16*, 167–176.
89. Camarero, J.A.; Mitchell, A.R. Synthesis of proteins by native chemical ligation using Fmoc-based chemistry. *Protein Pept. Lett.* **2005**, *12*, 723–728.
90. Clippingdale, A.B.; Barrow, C.J.; Wade, J.D. Peptide thioester preparation by Fmoc solid phase peptide synthesis for use in native chemical ligation. *J. Pept. Sci.* **2000**, *6*, 225–234.
91. Ali, A.M.; Taylor, S.D. Efficient solid-phase synthesis of sulfotyrosine peptides using a sulfate protecting-group strategy. *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 2024–2026.

92. Camarero, J.A.; Hackel, B.J.; de Yoreo, J.J.; Mitchell, A.R. Fmoc-based synthesis of peptide alpha-thioesters using an aryl hydrazine support. *J. Org. Chem.* **2004**, *69*, 4145–4151.
93. Woo, Y.-H.; Mitchell, A.; Camarero, J. The Use of Aryl Hydrazide Linkers for the Solid Phase Synthesis of Chemically Modified Peptides. *Int. J. Pept. Res. Ther.* **2007**, *13*, 181–190.
94. Haase, C.; Seitz, O. Internal Cysteine Accelerates Thioester-Based Peptide Ligation. *Eur. J. Org. Chem.* **2009**, *2009*, 2096–2101.
95. Camarero, J.A.; Pavel, J.; Muir, T.W. Chemical Synthesis of a Circular Protein Domain: Evidence for Folding-Assisted Cyclization. *Angew. Chem. Int. Ed. Engl.* **1998**, *37*, 347–349.
96. Schmoldt, H.U.; Wentzel, A.; Becker, S.; Kolmar, H. A fusion protein system for the recombinant production of short disulfide bond rich cystine knot peptides using barnase as a purification handle. *Protein Expr. Purif.* **2005**, *39*, 82–89.
97. Cascales, L.; Henriques, S.T.; Kerr, M.C.; Huang, Y.H.; Sweet, M.J.; Daly, N.L.; Craik, D.J. Identification and characterization of a new family of cell-penetrating peptides: Cyclic cell-penetrating peptides. *J. Biol. Chem.* **2011**, *286*, 36932–36943.
98. Lovelace, E.S.; Armishaw, C.J.; Colgrave, M.L.; Wahlstrom, M.E.; Alewood, P.F.; Daly, N.L.; Craik, D.J. Cyclic MrIA: A stable and potent cyclic conotoxin with a novel topological fold that targets the norepinephrine transporter. *J. Med. Chem.* **2006**, *49*, 6561–6568.
99. Edman, P. A method for the determination of amino acid sequence in peptides. *Arch. Biochem.* **1949**, *22*, 475.
100. Kratzner, R.; Debreczeni, J.E.; Pape, T.; Schneider, T.R.; Wentzel, A.; Kolmar, H.; Sheldrick, G.M.; Uson, I. Structure of Ecballium elaterium trypsin inhibitor II (EETI-II): A rigid molecular scaffold. *Acta Crystallogr. D Biol. Crystallogr.* **2005**, *61*, 1255–1262.
101. Craik, D.J.; Swedberg, J.E.; Mylne, J.S.; Cemazar, M. Cyclotides as a basis for drug design. *Expert Opin. Drug Discov.* **2012**, *7*, 179–194.
102. Contreras, J.; Elnagar, A.Y.; Hamm-Alvarez, S.F.; Camarero, J.A. Cellular uptake of cyclotide MCoTI-I follows multiple endocytic pathways. *J. Control. Release* **2011**, *155*, 134–143.
103. Reiss, S.; Sieber, M.; Oberle, V.; Wentzel, A.; Spangenberg, P.; Claus, R.; Kolmar, H.; Losche, W. Inhibition of platelet aggregation by grafting RGD and KGD sequences on the structural scaffold of small disulfide-rich proteins. *Platelets* **2006**, *17*, 153–157.
104. Getz, J.A.; Rice, J.J.; Daugherty, P.S. Protease-resistant peptide ligands from a knottin scaffold library. *ACS Chem. Biol.* **2011**, *6*, 837–844.
105. Tanimoto, H.; Yan, Y.; Clarke, J.; Korourian, S.; Shigemasa, K.; Parmley, T.H.; Parham, G.P.; O'Brien, T.J. Hepsin, A cell surface serine protease identified in hepatoma cells, Is overexpressed in ovarian cancer. *Cancer Res.* **1997**, *57*, 2884–2887.
106. Leytus, S.P.; Loeb, K.R.; Hagen, F.S.; Kurachi, K.; Davie, E.W. A novel trypsin-like serine protease (hepsin) with a putative transmembrane domain expressed by human liver and hepatoma cells. *Biochemistry* **1988**, *27*, 1067–1074.
107. Saleem, M.; Adhami, V.M.; Zhong, W.; Longley, B.J.; Lin, C.Y.; Dickson, R.B.; Reagan-Shaw, S.; Jarrard, D.F.; Mukhtar, H. A novel biomarker for staging human prostate adenocarcinoma: overexpression of matriptase with concomitant loss of its inhibitor, hepatocyte growth factor activator inhibitor-1. *Cancer Epidemiol. Biomark. Prev.* **2006**, *15*, 217–227.

108. Lin, C.Y.; Anders, J.; Johnson, M.; Dickson, R.B. Purification and characterization of a complex containing matriptase and a Kunitz-type serine protease inhibitor from human milk. *J. Biol. Chem.* **1999**, *274*, 18237–18242.
109. Jiang, L.; Miao, Z.; Kimura, R.H.; Silverman, A.P.; Ren, G.; Liu, H.; Lu, H.; Cochran, J.R.; Cheng, Z. 111In-labeled cystine-knot peptides based on the Agouti-related protein for targeting tumor angiogenesis. *J. Biomed. Biotechnol.* **2012**, *2012*, 368075.
110. Martin, L.; Stricher, F.; Misse, D.; Sironi, F.; Pugniere, M.; Barthe, P.; Prado-Gotor, R.; Freulon, I.; Magne, X.; Roumestand, C.; *et al*. Rational design of a CD4 mimic that inhibits HIV-1 entry and exposes cryptic neutralization epitopes. *Nat. Biotechnol.* **2003**, *21*, 71–76.
111. Stricher, F.; Huang, C.C.; Descours, A.; Duquesnoy, S.; Combes, O.; Decker, J.M.; Kwon, Y.D.; Lusso, P.; Shaw, G.M.; Vita, C.; *et al*. Combinatorial optimization of a CD4-mimetic miniprotein and cocrystal structures with HIV-1 gp120 envelope glycoprotein. *J. Mol. Biol.* **2008**, *382*, 510–524.
112. Vita, C.; Drakopoulou, E.; Vizzavona, J.; Rochette, S.; Martin, L.; Menez, A.; Roumestand, C.; Yang, Y.S.; Ylisastigui, L.; Benjouad, A.; *et al*. Rational engineering of a miniprotein that reproduces the core of the CD4 site interacting with HIV-1 envelope glycoprotein. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13091–13096.
113. Drakopoulou, E.; Vizzavona, J.; Vita, C. Engineering a CD4 mimetic inhibiting the binding of the human immunodeficiency virus-1 (HIV-1) envelope glycoprotein gp120 to human lymphocyte CD4 by the transfer of a CD4 functional site to a small natural scaffold. *Lett. Pept. Sci.* **1998**, *5*, 241–245.
114. Gunasekera, S.; Foley, F.M.; Clark, R.J.; Sando, L.; Fabri, L.J.; Craik, D.J.; Daly, N.L. Engineering stabilized vascular endothelial growth factor-A antagonists: Synthesis, Structural Characterization, And bioactivity of grafted analogues of cyclotides. *J. Med. Chem.* **2008**, *51*, 7697–7704.
115. Pranting, M.; Loov, C.; Burman, R.; Goransson, U.; Andersson, D.I. The cyclotide cycloviolacin O2 from Viola odorata has potent bactericidal activity against Gram-negative bacteria. *J. Antimicrob. Chemother.* **2010**, *65*, 1964–1971.

3 Experimenteller Teil

3.1 Oxidative Folding of Peptides with Cystine-knot Architechture: Kinetic studies and optimization of folding conditions.

Title

Oxidative Folding of Peptides with Cystine-Knot Architectures: Kinetic Studies and Optimization of Folding Conditions

Authors

Michael Reinwarth, Bernhard Glotzbach, Michael Tomaszowski, Sebastian Fabritz, Dr. Olga Avrutina* and Prof. Dr. Harald Kolmar*

Bibliographical Data

ChemBioChem, **2013**, 14(1): 137–146.

doi: 10.1002/cbic.201200604

Received: 19 September 2012 / Article first published online: 11 December 2012 / Issue published online: 23 December 2012

Contributions by Michael Reinwarth

- Synthesis of all reduced peptides
- Evaluation of different systems for oxidative folding including kinetic studies
- Execution of chromatographic and mass spectrometric experiments
- Realization of semi-preparative synthesis of oxidatively folded miniproteins for acitivity tests
- Creation of manuscript including all figures

DOI: 10.1002/cbic.201200604

Oxidative Folding of Peptides with Cystine-Knot Architectures: Kinetic Studies and Optimization of Folding Conditions

Michael Reinwarth, Bernhard Glotzbach, Michael Tomaszowski, Sebastian Fabritz, Olga Avrutina,* and Harald Kolmar*[a]

Bioactive peptides often contain several disulfide bonds that provide the main contribution to conformational rigidity and structural, thermal, or biological stability. Among them, cystine-knot peptides—commonly named "knottins"—make up a subclass with several thousand natural members. Hence, they are considered promising frameworks for peptide-based pharmaceuticals. Although cystine-knot peptides are available through chemical and recombinant synthetic routes, oxidative folding to afford the bioactive isomers still remains a crucial step. We therefore investigated the oxidative folding of ten protease-inhibiting peptides from two knottin families, as well as that of an HIV entry inhibitor and of aprotinin, under two conventional sets of folding conditions and by a newly developed procedure. Kinetic studies identified folding conditions that resulted in correctly folded miniproteins with high rates of conversion even for highly hydrophobic and aggregation-prone peptides in concentrated solutions.

Introduction

Natural polypeptides containing multiple disulfide bonds with mechanically interlocked topologies are found in diverse organisms: arthropoda, mollusca, porifera, vertebrata, fungi, and plantae.[1] These biomolecules, known as cystine-knot peptides or "knottins", usually possess peptide backbones of about 30 amino acid residues.[2] Being rather small, they nevertheless demonstrate the properties of full-sized proteins due to their unique architectures, each defined by three β-strands that are interconnected by three disulfide bonds, in which CysI of the sequence is connected to CysIV, CysII to CysV, and CysIII to CysVI (Figure 1).[2] The interlocked topology is formed by two disulfide bonds that, together with the connecting loops, form a ring through which the third disulfide bond is threaded. This results in a rigid structural core with exceptional thermal and biological stability. Knottins can be endowed with new bioactivities by variation of surface-exposed loops and are therefore considered to be promising peptide scaffolds for the generation of tailor-made bioactive compounds for various diagnostic and therapeutic applications.[3] Several miniproteins have already been applied as frameworks for the development of peptide-based pharmaceuticals, among them cyclotides, spider toxins, scyllatoxin, and squash trypsin inhibitors.[4] Knottins that act as, for example, potent inhibitors of human mast cell tryptase, an enzyme that plays a role in inflammatory diseases, especially in asthma, have been designed.[5] An engineered form of the human agouti-related protein (AgRP), a cystine-knot peptide with four disulfide bonds, was successfully used in radiolabeling for tumor imaging.[3b] Moreover, Ziconotide, a cystine-knot peptide from cone snails, has been on the market for treatment of severe and chronic pain since 2004.[6]

Thanks to their small sizes, knottins are readily accessible through solid-phase peptide synthesis.[2c] However, their chemical synthesis is often hampered by low folding yields due to the formation of incorrect intra- and intermolecular disulfide bonds.[2c] Although the investigative focus on the pathways of cystine knot formation is currently set on cyclotides—cyclic cystine-knot peptides in which the amino and the carboxy termini are linked by single peptide bonds—detailed studies on acyclic knottins have been carried out for only a few members of this large family.[7] In cases in which folding conditions have been systematically optimized for a particular knottin of interest, it has remained unclear whether these conditions are also applicable to other cystine-knot peptides.[2c,8]

In this study, we have investigated the folding of cysteine-rich peptides from different knottin families, as well as that of aprotinin, a three-disulfide trypsin inhibitor of the Kunitz-type family similar to basic pancreatic trypsin inhibitor (BPTI).[9] Peptides **1–8** (Figure 1 A) are based on the *Momordica cochinchinensis* trypsin inhibitor II (MCoTI-II) and therefore share the architecture of an inhibitor cystine knot (ICK) with the functional loop located between CysI and CysII.[10] During the last decade, ICK peptides have been intensively used as engineered frameworks in pharmaceutically inspired research.[1a,3c,d,5,11]

Recently reported three-disulfide peptides isolated from *Spinacia oleracea* seeds (SOTI I–III) also attracted our attention as

[a] *M. Reinwarth, B. Glotzbach, M. Tomaszowski, S. Fabritz, Dr. O. Avrutina, Prof. Dr. H. Kolmar*
Institute for Organic Chemistry and Biochemistry
Technische Universität Darmstadt
Petersenstr. 22, 64287 Darmstadt (Germany)
E-mail: Avrutina@Biochemie-TUD.de
Kolmar@Biochemie-TUD.de

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cbic.201200604.

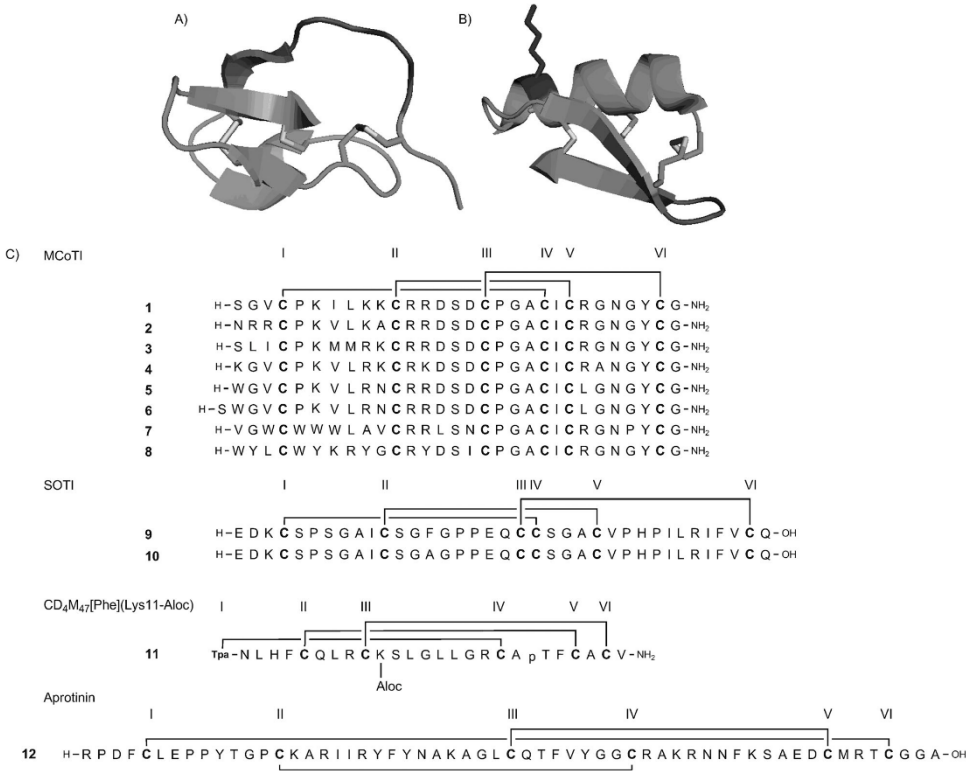


Figure 1. Overview of the miniproteins investigated in this work. A) Structure of an open-chain variant of MCoTI-II.[3a] B) Structure of CD4M47.[14b] C) Amino acid sequences of bioactive miniproteins presented in this work. The inhibitory loop can be seen top in (A) and bottom right in (B); Lys11 of CD4M47 is top left in (B); three disulfide bridges are shown in each. Tpa stands for thiopropiolic acid, p for D-proline, and Aloc for allyloxycarbonyl. Cysteines (and Tpa) are marked bold and numbered according to their appearance in the sequence. Cystine connectivities are indicated for each family.

potential scaffolds for the design of miniprotein libraries directed towards new protease binders against members of the type II serine protease family (compounds **9** and **10**).[1b,3b,5,12] Surprisingly, trypsin inhibitor SOTI I shows similarity not to any of the known plant protease inhibitors, but to a class of antimicrobial peptides from the seeds of *Mirabilis jalapa* with the inhibitory loop located between CysV and CysVI.[12b,13]

CD4M47$^{[Phe]}$ (**11**, Figure 1 B) was derived from scyllatoxin of *Leiurus quinquestriatus hebraeus*.[4d,14] This peptide has been optimized for binding to HIV-1 glycoprotein gp120 through several rounds of protein design and directed evolution, resulting in peptides with dissociation constants in the low nanomolar range towards the viral transmembrane protein.[4d,14a,b] Interestingly, this peptide is amino-terminally capped with dithiopropiolic acid (diTpa); this results in amino-terminal thiopropiolic acid (Tpa) that mimics homocysteine upon folding.

Aprotinin is a trypsin inhibitor obtained from bovine lung, and is identical to the more common bovine pancreatic trypsin inhibitor (BPTI), despite their different origins (bovine pancreas or lung, respectively). This inhibitor consists of 58 amino acids, among them six cysteine residues, and possesses a high proportion of basic amino acid residues; this makes the peptide a very potent inhibitor of trypsin.[9b,c,15] Folded BPTI contains three disulfides with a I–VI, II–IV, III–V connectivity that is different from that found in cystine-knot peptides **1**–**11**. Nevertheless, this peptide was also included in this study because in terms of oxidative folding and structural biology, BPTI is one of the most thoroughly studied trypsin inhibitors with clinical relevance. It reduces blood clotting through the inhibition of plasmin and until 2007 was sold under the registered trade name "Trasylol" by Bayer AG.[9c,15–16]

Although cystine-knot peptides are available through an abundant repertoire of peptide synthesis methods, oxidative folding still remains the crucial synthetic step.[2c,5,17] Very often, the formation of authentic disulfide connectivities requires a sophisticated regioselective strategy implying orthogonal thiol protection, polymer-supported oxidation, and isosteric disulfide replacements to ensure the correct crosslinking in the

presence of redox-favored species, or even deletion of a non-critical cystine pair or some combination of these methods.[2c,18] Moreover, folding yields are usually tightly correlated with high dilutions of the peptide, thus complicating workup of the folding mixtures.[2c] To date, no implementable generic approach to cystine bridging has been described.[2c]

Here we report the oxidative folding kinetics of a set of cystine-knot miniproteins, as well as the three-disulfide protease inhibitor aprotinin, and present a generic folding system that, in combination with ultrasonically assisted presolvation, allows one to obtain correctly disulfide-bonded knottins in high yields.

Results

Miniproteins of different origin were subjected to oxidative folding. Among them were the MCoTI-based open-chain protease inhibitors **1**–**8** (Figure 1), obtained from high-throughput screening experiments against matriptase or hepsin, which are known to be overexpressed in tumor cells, SOTI III and its Ala14 mutant (compounds **9** and **10**, respectively), and the scyllatoxin-derived CD 4 mimetic **11**, which contains a solubility-decreasing allyloxycarbonyl (Aloc) protecting group for downstream bioconjugation.[4d,13,14b,19] Although bioactive peptides of the same families have very similar backbone structures, differences in their individual sequences can result in remarkable differences in aggregation and folding behavior. The efficiency of our new folding system was determined by using reduced multicysteine peptide precursors that differed in their overall charges, polarities, and three-dimensional structures. Eight slightly basic cystine-knot peptides of the MCoTI family, some of them with large patches of hydrophobic or aromatic residues, two acidic knottins of the SOTI family, and one basic miniprotein derived from scyllatoxin, as well as the highly basic trypsin inhibitor aprotinin, were thus subjected to oxidative folding.

Optimization of oxidative folding conditions

Folding kinetics were followed by reversed-phase HPLC (Figures 2, 3, 4, and 5), as exemplified by folding of knottins **2** and **9**. Different oxidative folding conditions were applied. Air oxygen, as well as mixtures of reduced and oxidized glutathione or of DMSO/TFE/acetonitrile/Gu·HCl (DTAG) were used to establish redox conditions that drive oxidative folding.

Because the product yields were rather low when the previously reported standard air oxygen or the glutathione redox pair in aqueous buffer were used (Table 1), folding conditions were further optimized with use of DMSO as oxidant and co-solvent.[2c,4a,8c,10] Variation of DMSO concentration revealed that folding in 20% DMSO resulted in reaction rates equal to those of the glutathione system, but at the cost of the formation of side products. Higher concentrations of DMSO (up to 50%) correlated with better solubility and even faster reaction rates, but gave large amounts of non-reducible byproducts (data not shown). Use of less DMSO (down to 5%) led to reduced formation of side products but to longer reaction times and decreased solubility; in some cases even aggregation of peptides and precipitation was observed (data not shown). From the various folding experiments it was concluded that a mixture of organic solvents containing 10% acetonitrile, 10% DMSO as an oxidation agent and solvent, 10% trifluoroethanol (TFE) as secondary structure stabilizing and solvating adjuvant, and 1 M

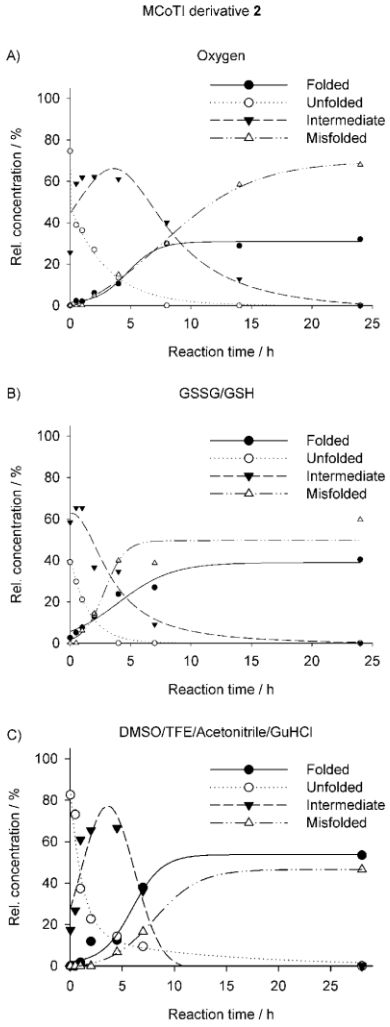


Figure 2. Kinetic investigation of MCoTI derivative **2** under different sets of folding conditions. A) Stirring in ammonium acetate buffer (pH 9.1) containing acetonitrile (10%). B) Acetonitrile (10%), oxidized/reduced glutathione (5 and 10 mM, respectively) in ammonium acetate buffer (pH 9.1). C) Acetonitrile (10%), DMSO (10%), TFE (10%), and guanidinium chloride (1 M) in aqueous phosphate buffer (pH 7.0). Data points were obtained by RP-HPLC analysis of the folding mixture in question at different time points. Relative concentrations were determined by peak integration.

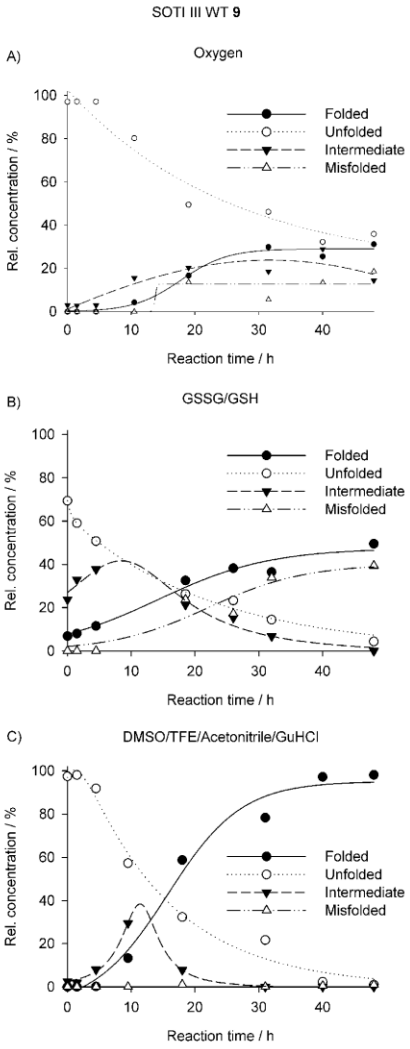


Figure 3. Kinetic investigation of SOTI derivative **9** under different sets of folding conditions. A) Stirring in ammonium acetate buffer (pH 9.1) containing acetonitrile (10%). B) Acetonitrile (10%), oxidized/reduced glutathione (5 and 10 mM, respectively) in ammonium acetate buffer (pH 9.1). C) DMSO (10%), acetonitrile (10%), TFE (10%), guanidinium chloride (1 M) in aqueous phosphate buffer (pH 7.0). Data points were obtained by RP-HPLC analysis of the folding mixture in question at different time points. Relative concentrations were determined by peak integration.

aqueous guanidinium hydrochloride (Gu·HCl) as aggregation-disrupting and solvating additive at pH 7 gave the highest yields relative to commonly used folding conditions both on analytical and on multimilligram scales (Figure 4, Table 1).[3b,8c,20] Except for the folding of **12**, each experiment was conducted with a peptide concentration of 10 mg mL^{-1}.

Table 1. Comparison of folding yields after 24 h under different sets of oxidative conditions at peptide concentrations of 10 mg mL^{-1} in each experiment, except in the case of **12**, which was folded at 5 mg mL^{-1}.

	HPLC conversion [%]			Yield[d]
Mini-protein	1	2	3	[%]
1	17.5	22.6	49.1	22.0
2	32	40.4	53.4	19.1
3	9.7	16.2	80.6	27.8
4	0	20.2	34.9	18.2
5	0[a,b]	0[a,b]	72.1[b]	14.1[b]
6	41.1[b]	43.1[b]	50.5[b]	11.5[b]
7	0[a,b]	1.4[a,b]	42.5[b]	9.6[b]
8	0[a,b]	0[a,b]	48.7[b]	9.5[b]
9	31.1[c]	49.6[c]	98.1[c]	23.8[c]
10	42.7[c]	59.8[c]	85[c]	16.2[c]
11	0[a]	0[a]	54.4	7.0
12	0[a,c]	48.7[a,c]	72.4[a,c]	13.3[c]

Conditions: 1: air oxygen; 10% MeCN in 200 mM NH_4OAc, pH 9.1; 2: GSSG/GSH (5/10 mM) in 10% MeCN in 200 mM NH_4OAc, pH 9.1; 3: DMSO, TFE, MeCN (10% each) in 1 M Gu·HCl in 50 mM Na_2HPO_4, pH 7. [a] Aggregation as a white precipitate observed. [b] Reaction time of seven days. [c] Reaction time of two days. [d] On semipreparative scale with DMSO/TFE/Gu·HCl; yields refer to crude linear precursor.

The folding process was usually completed within 24 h, although in some difficult cases complete conversion took several days. Surprisingly, suspension of the peptide in acetonitrile

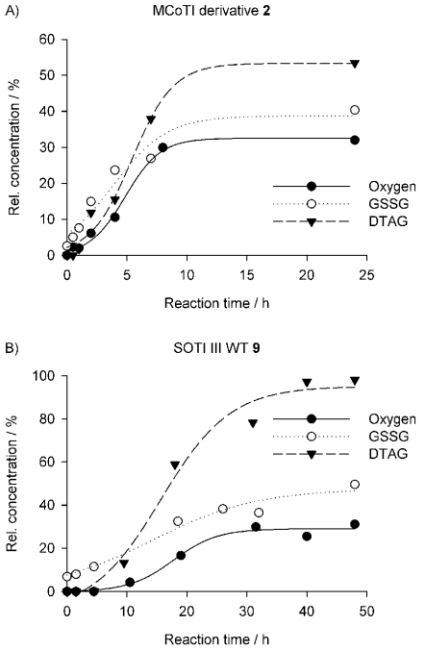


Figure 4. Comparison of product formation under the different sets of folding conditions. A) MCoTI derivative **2**. B) SOTI III WT (**9**).

in an ultrasonic bath was found to be a crucial step to avoid peptide aggregation in the folding process, especially during experiments on semipreparative scales; this is likely due to solvation of aggregation-prone regions. In particular, peptide **7**, with its unusually large number of tryptophan residues, was found to be very prone to aggregation in aqueous buffer systems. This peptide was obtained from screening experiments as an inhibitor (most likely allosteric) of the serine protease hepsin. Nevertheless, for aprotinin (**12**), with its numerous patches of basic amino acids, the DTAG folding conditions had to be slightly modified due to the fact that reduced aprotinin is almost completely insoluble in neutral or basic aqueous folding buffer (Table 1, Figure S40 in the Supporting Information). In this case, the peptide was predissolved in more concentrated (2.4-fold) DTAG, followed by dilution to 5 mg mL^{-1} final concentration.

Comparison of folding kinetics in DTAG folding buffer

Figure 5 shows RP-HPLC traces of folding reactions for three peptides, each a representative example of a knottin family, in the optimized DTAG-based folding buffer. Interestingly, folding intermediates of trypsin inhibitors **2** and **9** strongly differ with respect to their hydrophobicity. For MCoTI derivative **2** the observed folding intermediate—most likely the CysI–CysIV, CysII–CysV two-disulfide version—displays intermediate hydrophobicity relative to the native and the unfolded state (Figure 5, left).[2c,7b,21] During the folding of SOTI, an intermediate that displayed considerably higher hydrophobicity than the peptide in its unfolded state was accumulated (Figure 5, middle). The nature of this intermediate still has to be elucidated.

Biological activities of folded peptides

Although different disulfide connectivities can be distinguished by comparison of RP-HPLC retention times on a C18 column, it remained to be investigated whether the obtained oxidized species possessed the correctly knotted pattern.[2c,8c,22] For knottins that act as protease inhibitors it is known that inhibitory activities are strictly dependent on particular disulfide bond topologies.[5,23]

Wild-type sequences of MCoTI and SOTI with known K_i values against trypsin were used as references for determination of the correct disulfide patterns. Similarity between the K_i values obtained in this work and those previously reported indicated the formation of correct disulfide patterns.[10,12b,13] Additionally, the grafted miniproteins displayed behavior analogous to that of their wild-type counterparts upon HPLC monitoring during folding. For miniproteins **7** and **8** the inhibition constants were determined not only for the final products but also for the major folding byproducts, revealing K_i values in the micro- to millimolar range, hence clarifying their misfolded natures (data not shown). A combination of both methods is thus a valuable tool to test native disulfide bond formation. All 11 protease inhibitors examined in this study showed inhibition constants in the nanomolar range against their corresponding target proteases (Table 2). Correct folding of **11** was verified by CD spectroscopy (Figure S37).

N-terminal aldehyde installation in a one-pot reaction

Peptides **1**, **3**, and **6** each contain an N-terminal serine residue that can be used for selective periodate oxidation to generate

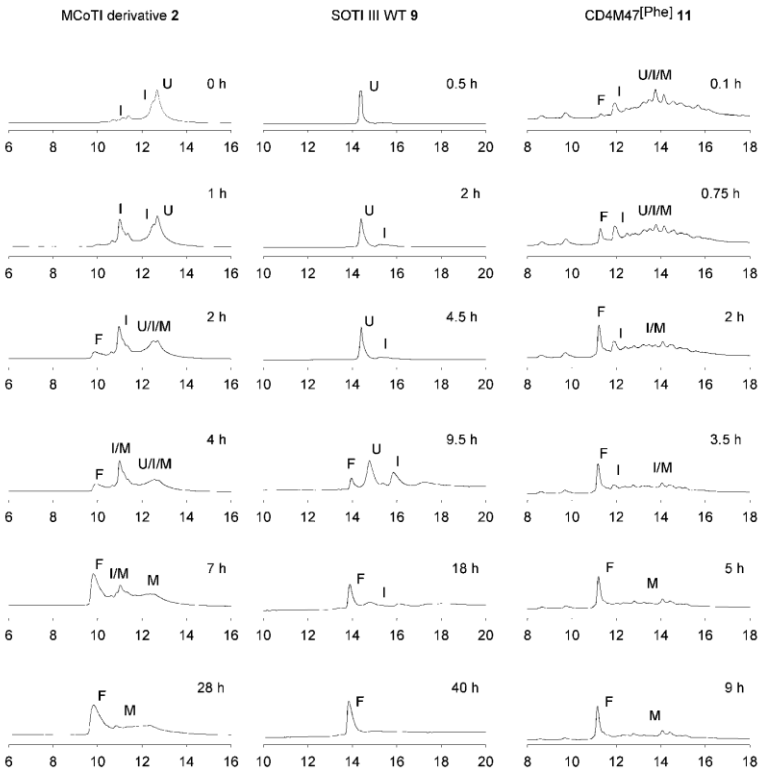


Figure 5. HPLC monitoring of folding of **2**, **9**, and **11** in DTAG folding buffer at different reaction times. F: folded products, M: misfolded byproducts, I: folding intermediates, U: unfolded precursor.

Table 2. Average inhibition of folded miniprotein variants against trypsin or hepsin.

Miniprotein	K_i [nM]	Miniprotein	K_i [nM]
1	0.15	**6** (glyoxylyl)	24
1 (glyoxylyl)	0.5	**7**	131[a]
2	2.3	**8**	100[a]
3	1.2	**9**	58
3 (glyoxylyl)	589	**10**	194
4	2.3	**12** (lung isolate)	0.11
5	2.2	**12** (refolded)	0.18
6	32		

[a] Determined against hepsin as a double determination.

a glyoxylyl moiety for subsequent site-specific conjugation as, for example, a fluorescent label or a radiolabel.[24] Addition of 10 equiv periodate to the DTAG mixture containing **1** or **6** resulted in complete conversion to the desired glyoxal or its corresponding hydrated form (Figure 6).[25] Periodate oxidation of MCoTI variant **3** also resulted in the generation of an N-terminal aldehyde with DTAG, but side reactions due to an unspecified oxidation of the vicinal methionine residues were observed and dramatically decreased inhibitory efficiency (Table 2). Under periodate-supplemented glutathione-based folding conditions, less or no N-terminal conversion (peptides **1**, **3**) or precipitation of peptide (**6**) was observed. For the oxidation of **1** and **3** increased amounts of sodium periodate (up to 50 equiv, based on peptide) caused no improvement and no oxidized peptide was noticeable. In conclusion, oxidation of amino-terminal serine in high yield and purity can easily be accomplished in a stepwise (serine oxidation after folding) one-pot reaction.

Discussion

In recent years, knottins have frequently been used as rigid and stable structural scaffolds for the introduction of new biological activities by varying the lengths and sequences of loops and also by substituting the framework residues.[5, 14b, c, 26] Many natural inhibitor cystine-knot peptides were found to be aggregation-prone, particularly in their reduced forms. For engineered variants this tendency can increase with increasing amounts of amino acid variation relative to the wild types. In this study, folding of cysteine-rich peptides derived from the trypsin inhibitors MCoTI II and SOTI, as well as a scyllatoxin derivative, was investigated. Variants displaying minor polarity changes within their sequences (**2**, **3**, **4**, **6**) demonstrated folding behavior similar to that of the wild type (**1**), although significant polarity changes—exchange of arginine for leucine or more hydrophobic residues (**5**, **7**, **8**), for example—resulted in highly aggregation-prone peptides hardly soluble in aqueous media. Yields of correctly folded knottins correlated with the solubilities of the corresponding linear precursors, as is highlighted by comparison of peptides **5** and **6**, in which N-terminal elongation with serine resulted in considerable increases of correctly folded miniprotein in all three folding systems (Figure 1, Table 1). Alteration of the pH in the glutathione-based systems revealed no relevant changes in conversion to the correctly folded products or their yields, as determined representatively for knottin **2** (data not shown).

Although the explicit folding mechanism of SOTI could not be investigated in the context of this project, RP-HPLC indicates a kinetically trapped intermediate of relatively hydrophobic character (Figure 5). This intermediate might involve an established CysII–CysV and CysIII–CysVI connection, because this is the common energetic trap for peptides of the inhibitor cystine-knot family.[2c, 7b, 27] In contrast to what might be intuitively expected, a longer RP-HPLC retention time than for the unfolded form was observed for this intermediate with hydrophobic residues present on the outer shell. The rate-limiting final folding step requires compaction and the reorientation of the hydrophobic and hydrophilic residues towards the inner core or the outer shell, respectively. The most demanding knottin in this series with respect to folding is the HIV entry inhibitor **11**, because the presence of an Aloc protecting group at Lys11 dramatically reduces solubility in aqueous buffers, so the unfolded peptide could not be dissolved in the ammonium acetate buffer and air oxygen or a glutathione mixture could not be used as ox-

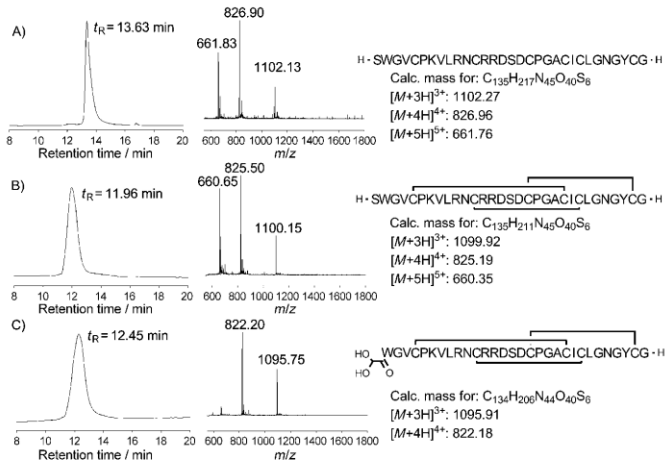


Figure 6. One-pot folding and oxidation of MCoTI variant **6**. A) HPLC trace and ESI-MS of the linear precursor of MCoTI derivative **6**. B) HPLC trace and ESI-MS of folded miniprotein **6**. C) HPLC trace and ESI-MS of folded and N-terminally oxidized MCoTI derivative **6**. t_R: RP-HPLC retention time.

idants. The fast folding in the DTAG folding cocktail—within five hours—indicates that this mixture is particularly useful for oxidative folding of hydrophobic and aggregation-prone peptides (Figure 5). For aprotinin, which is usually produced recombinantly fused to a polar propeptide that assists folding, the experimental procedure with the DTAG system had to be altered because of the high content of basic amino acid residues in the sequence that resulted in reduced solubility at pH 7 and pH 9.1 (see the Experimental Section).[28] As a consequence, the reduced peptide was directly dissolved in a 2.4-fold concentrated DTAG mixture followed by dilution with phosphate buffer pH 7.0 to 5 mg mL^{-1} and 2.5 M Gu·HCl, resulting in folding conditions nearly identical to those used for all other peptides.

Oxidation of cysteine residues is usually mediated by air, cystines, or dimethyl sulfoxide.[29] In the case of air oxidation of knottins, slow reaction rates and requirements for high dilution, especially for aggregation-prone peptides, are two major problems.[2c] Use of stronger oxidizing agents can cause significant side reactions through incorrect bridging or overoxidation of sulfur-containing residues. Both undesirable processes lead to the loss of particular architecture and biological activity. The necessity to conduct oxidative folding of hydrophobic peptides under conditions of very high dilution significantly complicates their post-synthetic purification, so the use of a polar organic co-solvent such as TFE, which solubilizes hydrophobic side chains and also assists in the formation and stabilization of α-helices, especially in combination with a mild oxidizing agent as DMSO, is often an alternative approach.[30] Additionally, some structure-disrupting blocks—proline or proline-like surrogates, for example—are known to reduce aggregation and have been used as additives in several folding buffers.[31] Other systems involving ionic liquids and platinum-based adjuvants are quite laborious, impractical, or narrow in scope for common use.[32]

The DTAG knottin folding cocktail developed in this study contains DMSO as oxidizing agent, TFE as cosolvent, and acetonitrile as solvent, together with aqueous guanidinium hydrochloride as chaotropic agent. The absence of guanidinium hydrochloride frequently resulted in detectable aggregation and precipitation that was not observed in the presence of 1 M guanidinium hydrochloride. At this concentration the oxidation process itself can be expected not to be influenced negatively by this denaturing agent, whereas disfavored peptide aggregation, particularly at relatively high peptide concentrations in the DTAG folding cocktail (usually 10 mg mL^{-1}), can be suppressed until the peptide becomes more hydrophilic and less aggregation-prone by compaction through oxidative folding.

The aggregation tendencies of the peptides, which are closely correlated with but not equal to their polarities, were found to be a key factor upon oxidative folding. Aggregation under oxidative conditions most likely results in the formation of disulfide-crosslinked polymers that are very hard to reduce or to redissolve, even by treatment with reducing agents such as dithiothreitol at high concentrations or with Zn as reducing agent in TFA/H_2O.[33] Ultrasonic-assisted presolvation in acetonitrile in combination with organic solvents and chaotropic salts is likely to reduce the aggregation tendency during the oxidation reaction.

Oxidative folding of multidisulfide-containing peptides by, for example, air oxygen without any reducing agent often results to large extent in misfolded byproducts, because cysteine thiols, once incorrectly bridged, lack the capability for spontaneous reduction.[2c] Addition of a redox pair such as ox./red. glutathione or cystine/cysteine to the folding mixture usually increases yields and purities of the correctly folded products.[2c, 20, 26, 34] Interestingly, the DTAG mixture that provided the highest folding yields did not contain any reducing agent. However, inaccurately connected cysteine residues have low thermodynamic barriers to their reduction, so unpaired thiols in the peptide backbone might serve as reducing agents in a rearrangement reaction; the relatively low reduction potential of DMSO/DMS (160 mV) relative to GSSG/GSH (240 mV) or ($^1/_2O_2 + H_2$)/H_2O (410 mV) at pH 7 likely provides a favorable redox environment.[11, 35] The side products of DMSO-based folding might therefore act as mild oxidants, thereby allowing for rearrangements, whereas side products of oxygen- or glutathione-based folding are often trapped in energetically uncorrectable states (Table 1).

Very often, bioactive peptides need incorporation of functional moieties with orthogonal reactivity to act as coupling sites for further modifications, such as for introduction of labeling compounds, conjugation with biomolecules, grafting onto surfaces, ligation with hybrid inorganic targets, or multivalent presentation on different scaffolds for enhanced bioactivity and bioavailability.[36] Although a variety of orthogonal active groups and conjugation techniques are available for peptidic molecules, the majority of ligation methods are not suitable for reactions involving cysteine-rich peptides.[37] Chemoselective oxidation of N-terminal serine, however, yields a reactive glyoxylyl functionality that can easily react with aminooxy, hydrazine, hydrazide, or thiosemicarbazide groups to form stable covalent bonds under mild, disulfide-friendly conditions.[38] Although periodate oxidation is a fully elaborated method, it needs an additional purification step, thus decreasing the final yield of the desired peptide.[37c, 38a]

An additional benefit of the DTAG folding cocktail is the option to add sodium periodate to the reaction mixture directly for oxidation of an N-terminal serine unit to an aldehyde without affecting already bridged cystine pairs. Unwanted oxidation of methionine residues can occur, however (Figure S12), so periodate concentration and reaction time require careful adjustment.[39] Yields of oxidation product similar to those achieved with commonly used aqueous solvents were obtained for DTAG mixtures.[25a] Notably, subsequent in situ oxidation of the aldehyde group to the acid was not observed.[40]

Conclusions

Oxidative folding of cysteine-rich peptides is still a challenging matter because individually tuned folding conditions are often necessary. Particularly demanding is the formation of three disulfides in small peptides possessing cystine-knot architectures.[2c] These kinetic studies have identified folding conditions

that have resulted in correctly folded products even in cases of hydrophobic peptides containing multiple thiol units with strong aggregation tendencies in concentrated solutions (10 mg mL^{-1}) with excellent conversion rates. As demonstrated for aprotinin, this method might not be limited to the oxidative folding of peptides of the cystine-knot family but might also provide a starting point for optimizing folding conditions of other peptides with high cysteine contents and different disulfide patterns.

Experimental Section

All reagents were used without further purification as received from commercial sources. Fmoc-protected amino acids and activators for solid-phase peptide synthesis (SPPS) were obtained from IRIS Biotech (Marktredwitz, Germany); common reagents and solvents were purchased from Sigma–Aldrich, Carl Roth, or Fisher Scientific.

Synthesis of linear precursors: Peptides were assembled by standard Fmoc SPPS chemistry with a fully automated microwave-assisted CEM Liberty peptide synthesizer. Peptide amides were generated with a ChemMatrix Fmoc-Rink amide resin (Agilent Technologies, Böblingen, Germany). Peptide acids were synthesized on Fmoc-Gln preloaded TentaGel resin (Rapp Polymere, Tübingen, Germany). Amino-terminal capping of the HIV entry inhibitor **11** with Tpa was conducted by a standard coupling strategy with Tpa (4 equiv), 2-(1*H*-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphonate (HBTU; 3.9 equiv) and *N*,*N*-diisopropylethylamine (DIEA; 8 equiv). Cleavage from the solid support was generally achieved with TFA (94%, *v*/*v*), H_2O (2%, *v*/*v*), dithiothreitol (DTT, 2%, *w*/*v*), triethylsilane (1%, *v*/*v*), and anisole (1%, *v*/*v*). After precipitation of the crude peptide in methyl *tert*-butyl ether, the pellet was dissolved in a mixture of acetonitrile and water (2:8, *v*/*v*), lyophilized, and analyzed by RP-HPLC and LC-ESI-MS.

RP-HPLC, LC-ESI-MS, and CD spectroscopy: Analytical RP-HPLC was performed with a Varian LC 920 system and a Phenomenex Synergi 4μ Hydro-RP 80 Å (250×4.6 mm, 4 μm) column with application of linear gradients of acetonitrile at a flow rate of 1 mL min^{-1}. Semipreparative RP-HPLC was performed with a Varian LC 940 system and an axia-packed Phenomenex Luna C18 (250×21.2 mm, 5 μm, 100 Å) column with application of linear acetonitrile gradients at a flow rate of 18 mL min^{-1}. Isocratic elution [10% eluent B for 2 min (analytical scale) or 5 min (semipreparative scale)] was followed by a linear gradient of 10→60% B or 10→80% B, respectively, over 20 min.

LC-MS was performed with a Shimadzu LC-MS 2020 instrument and a Phenomenex Jupiter C4 (50×1 mm, 5 μm, 300 Å) column with linear acetonitrile gradients and a flow rate of 0.2 mL min^{-1} (see Figures S1–S40). Isocratic elution (2% eluent B for 2 min) was followed by a linear gradient of 2→100% B over 10 min.

CD spectroscopy was performed as previously reported.[14b] Peptide **11** was thus dissolved in aqueous Na_2HPO_4 (pH 7, 2 mM) to a final concentration of 50 μM. The resulting spectrum (Figure S37) was obtained through accumulation of 100 spectra with use of a 0.1 mm quartz cuvette and 0.5 nm steps.

Integration of HPLC chromatograms was performed with Varian Galaxie chromatography software (Version 1.9.302.952). The graphic display was achieved with the "dynamic plot" function of Sigma Plot 11.

Stepwise one-pot folding and amino-terminal oxidation of cystine-knot peptides: Oxidative folding on analytic scale was performed by three different approaches. In each case the appropriate purified and lyophilized peptide (1 mg) was suspended in acetonitrile (10 μL) and treated in an ultrasonic bath for 5 min. Afterwards, the appropriate folding mixture (90 μL) was added: mixture A was composed of aqueous ammonium acetate (200 mM, pH 9.1), mixture B of oxidized/reduced glutathione (5/10 mM) in aqueous ammonium acetate (200 mM, pH 9.1), and mixture C of DMSO (10%, *v*/*v*), TFE (10%, *v*/*v*), and Gu·HCl (1 M) in aqueous sodium phosphate buffer (pH 7, 50 mM). Oxidative folding was usually complete within 24 h at room temperature; slow-folding peptides were kept for two or seven days (Table 1). Reaction was stopped through direct injection into analytical HPLC. On a semipreparative scale, folding of mixture C was applied as described with a final concentration of the crude peptide of 10 mg mL^{-1}, while the ratios of the individual folding assistants used were kept equal in the system.

Oxidation of N-terminal serine was performed by direct addition of $NaIO_4$ (10 equiv with respect to the peptide) to the folding mixture. Reaction was terminated by direct injection into the HPLC or LC-ESI-MS system for purification or analysis, respectively.

Unfolding and refolding of aprotinin: For reductive unfolding, aprotinin (Carl Roth) was dissolved in Gu·HCl (6 M). After addition of DTT (10 equiv) the reaction mixture was kept at ambient temperature for 2 h. It was directly injected into semipreparative HPLC.

Oxidative refolding of the reduced precursor was performed as for the cystine-knot peptides with slight modifications of the DTAG-based method. The reduced peptide was suspended in acetonitrile and treated in an ultrasonic bath for 5 min. Afterwards, the prefolding mixture containing DMSO (25%, *v*/*v*), TFE (25%, *v*/*v*), and Gu·HCl (6 M) in aqueous sodium phosphate buffer (pH 7, 50 mM) was added at 12.5 mg mL^{-1}. After additional treatment in an ultrasonic bath until the peptide was completely dissolved, the mixture was diluted 2.5 times with aqueous sodium phosphate buffer (pH 7, 50 mM). Oxidative folding was complete within 48 h at room temperature; formation of a small amount of insoluble white precipitate was observed.

Inhibition assays: The substrate-independent inhibition constants were determined as described previously.[10,41] Generally, data were collected in triplicate with a Tecan Genios ELISA reader. In the case of hepsin, measurements were performed in duplicate. The normalized residual proteolytic activities (v/v_0) of trypsin (Sigma–Aldrich), or hepsin (1 nM) towards the chromogenic substrates Boc-QAR-pNA (250 μM) or Boc-QAR-AMC (250 μM, Bachem, Weil am Rhein, Germany), respectively, in the presence of inhibitors at different concentrations were determined after 30 min by monitoring the absorbance at 405 or 465 nm for 30 min. Apparent inhibition constants (K_i^{app}) were calculated by fitting the data with use of the Morrison equation [Eq (1)] for tight-binding inhibitors with the Marquardt–Levenberg algorithm of SigmaPlot 11.[42]

$$\frac{v}{v_0} = 1 - \frac{(E_0 + I_0 + K_i^{app}) - \sqrt{E_0 + I_0 + K_0 - 4E_0I_0}}{2E_0} \quad (1)$$

Substrate-independent inhibition constants K_i were calculated from K_i^{app} and K_M of the enzyme according to Equation (2).

$$K_i = \frac{K_i^{app}}{1 + [S]/K_M} \quad (2)$$

Acknowledgements

This work was supported in part by the Deutsche Forschungsgemeinschaft through SPP1623 grant KO 1390/10-1 and by BMBF project BioTrace.

Keywords: aprotinin · cystine knots · inhibitors · oxidative folding · peptides

[1] a) L. Chiche, A. Heitz, J. C. Gelly, J. Gracy, P. T. Chau, P. T. Ha, J. F. Hernandez, D. Le-Nguyen, *Curr. Protein Pept. Sci.* **2004**, *5*, 341 – 349; b) H. Kolmar, *Curr. Opin. Pharmacol.* **2009**, *9*, 608 – 614.

[2] a) J. Gracy, D. Le-Nguyen, J. C. Gelly, Q. Kaas, A. Heitz, L. Chiche, *Nucleic Acids Res.* **2008**, *36*, D314 – D319; b) J. C. Gelly, J. Gracy, Q. Kaas, D. Le-Nguyen, A. Heitz, L. Chiche, *Nucleic Acids Res.* **2004**, *32*, D156 – 159; c) M. Reinwarth, D. Nasu, H. Kolmar, O. Avrutina, *Molecules* **2012**, *17*, 12533 – 12552.

[3] a) A. Heitz, O. Avrutina, D. Le-Nguyen, U. Diederichsen, J. F. Hernandez, J. Gracy, H. Kolmar, L. Chiche, *BMC Struct. Biol.* **2008**, *8*, 54; b) R. H. Kimura, Z. Cheng, S. S. Gambhir, J. R. Cochran, *Cancer Res.* **2009**, *69*, 2435 – 2442; c) M. Werle, K. Kafedjiiski, H. Kolmar, A. Bernkop-Schnurch, *Int. J. Pharm.* **2007**, *332*, 72 – 79; d) M. Werle, T. Schmitz, H. L. Huang, A. Wentzel, H. Kolmar, A. Bernkop-Schnurch, *J. Drug Targeting* **2006**, *14*, 137 – 146.

[4] a) O. Avrutina, H. U. Schmoldt, D. Gabrijelcic-Geiger, A. Wentzel, H. Frauendorf, C. P. Sommerhoff, U. Diederichsen, H. Kolmar, *ChemBioChem* **2008**, *9*, 33 – 37; b) J. I. Fletcher, B. E. Chapman, J. P. Mackay, M. E. Howden, G. F. King, *Structure* **1997**, *5*, 1525 – 1535; c) S. Gunasekera, F. M. Foley, R. J. Clark, L. Sando, L. J. Fabri, D. J. Craik, N. L. Daly, *J. Med. Chem.* **2008**, *51*, 7697 – 7704; d) C. Vita, E. Drakopoulou, J. Vizzavona, S. Rochette, L. Martin, A. Menez, C. Roumestand, Y. S. Yang, L. Ylisastigui, A. Benjouad, J. C. Gluckman, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13091 – 13096.

[5] C. P. Sommerhoff, O. Avrutina, H. U. Schmoldt, D. Gabrijelcic-Geiger, U. Diederichsen, H. Kolmar, *J. Mol. Biol.* **2010**, *395*, 167 – 175.

[6] U. Klotz, *Int. J. Clin. Pharmacol. Ther.* **2006**, *44*, 478 – 483.

[7] a) M. Cemazar, C. W. Gruber, D. J. Craik, *Antioxid. Redox Signaling* **2008**, *10*, 103 – 111; b) D. Le-Nguyen, A. Heitz, L. Chiche, M. el Hajji, B. Castro, *Protein Sci.* **1993**, *2*, 165 – 174; c) S. Gunasekera, N. L. Daly, R. J. Clark, D. J. Craik, *Antioxid. Redox Signaling* **2009**, *11*, 971 – 980.

[8] a) T. L. Aboye, R. J. Clark, R. Burman, M. B. Roig, D. J. Craik, U. Goransson, *Antioxid. Redox Signaling* **2011**, *14*, 77 – 86; b) B. R. Green, G. Bulaj, *Protein Pept. Lett.* **2006**, *13*, 67 – 70; c) M. Price-Carter, W. R. Gray, D. P. Goldenberg, *Biochemistry* **1996**, *35*, 15537 – 15546.

[9] a) D. P. Goldenberg, *Trends Biochem. Sci.* **1992**, *17*, 257 – 261; b) H. Fritz, G. Wunderer, *Arzneim.-Forsch.* **1983**, *33*, 479 – 494; c) P. Ascenzi, A. Bocedi, M. Bolognesi, A. Spallarossa, M. Coletta, R. De Cristofaro, E. Menegatti, *Curr. Protein Pept. Sci.* **2003**, *4*, 231 – 251.

[10] O. Avrutina, H. U. Schmoldt, D. Gabrijelcic-Geiger, D. Le Nguyen, C. P. Sommerhoff, U. Diederichsen, H. Kolmar, *Biol. Chem.* **2005**, *386*, 1301 – 1306.

[11] F. W. Sweat, W. W. Epstein, *J. Org. Chem.* **1967**, *32*, 835 – 838.

[12] a) K. Jagadish, J. A. Camarero, *Biopolymers* **2010**, *94*, 611 – 616; b) B. Glotzbach, S. Schmelz, M. Reinwarth, A. Christmann, D. W. Heinz, H. Kolmar, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2012**, *in press*, DOI:10.1107/S0907444912043880.

[13] J. Kowalska, K. Pszczola, A. Wilimowska-Pelc, I. Lorenc-Kubis, E. Zuziak, M. Lugowski, A. Legowska, A. Kwiatkowska, M. Sleszynska, A. Lesner, A. Walewska, E. Zablotna, K. Rolka, T. Wilusz, *Phytochemistry* **2007**, *68*, 1487 – 1496.

[14] a) L. Martin, F. Stricher, D. Misse, F. Sironi, M. Pugniere, P. Barthe, R. Prado-Gotor, I. Freulon, X. Magne, C. Roumestand, A. Menez, P. Lusso, F. Veas, C. Vita, *Nat. Biotechnol.* **2003**, *21*, 71 – 76; b) F. Stricher, C. C. Huang, A. Descours, S. Duquesnoy, O. Combes, J. M. Decker, Y. D. Kwon, P. Lusso, G. M. Shaw, C. Vita, P. D. Kwong, L. Martin, *J. Mol. Biol.* **2008**, *382*, 510 – 524; c) E. Drakopoulou, J. Vizzavona, C. Vita, *Lett. Pept. Sci.* **1998**, *5*, 241 – 245.

[15] B. Kassell, M. Radicevic, M. J. Ansfield, M. Laskowski, Sr., *Biochem. Biophys. Res. Commun.* **1965**, *18*, 255 – 258.

[16] a) B. Kassell, M. Laskowski, Sr., *Biochem. Biophys. Res. Commun.* **1965**, *20*, 463 – 468; b) N. J. Darby, P. E. Morin, G. Talbo, T. E. Creighton, *J. Mol. Biol.* **1995**, *249*, 463 – 477; c) D. J. States, C. M. Dobson, M. Karplus, T. E. Creighton, *Nature* **1980**, *286*, 630 – 632; d) J. S. Weissman, P. S. Kim, *Science* **1991**, *253*, 1386 – 1393; e) A. Tuffs, *Br. Med. J.* **2007**, *335*, 1015; f) A. Deanda, Jr., B. D. Spiess, *J. Thorac. Cardiovasc. Surg.* **2012**, *144*, 998 – 1002.

[17] P. Da Silva, A. Strzepa, L. Jouvensal, I. Rahioui, F. Gressent, A. F. Delmas, *Biopolymers* **2009**, *92*, 436 – 444.

[18] a) M. Husbyn, A. Cuthbertson, *J. Pept. Res.* **2002**, *60*, 121 – 127; b) F. Shen, Z. P. Zhang, J. B. Li, Y. Lin, L. Liu, *Org. Lett.* **2011**, *13*, 568 – 571; c) K. Darlak, D. Wiegandt Long, A. Czerwinski, M. Darlak, F. Valenzuela, A. F. Spatola, G. Barany, *J. Pept. Res.* **2004**, *63*, 303 – 312; d) T. S. Han, M. M. Zhang, K. H. Gowd, A. Walewska, D. Yoshikami, B. M. Olivera, G. Bulaj, *ACS Med. Chem. Lett.* **2010**, *1*, 140 – 144; e) A. M. Steiner, G. Bulaj, *J. Pept. Sci.* **2011**, *17*, 1 – 7.

[19] a) C. Y. Lin, J. Anders, M. Johnson, Q. A. Sang, R. B. Dickson, *J. Biol. Chem.* **1999**, *274*, 18231 – 18236; b) K. List, *Future Oncol.* **2009**, *5*, 97 – 104; c) K. Uhland, *Cell. Mol. Life Sci.* **2006**, *63*, 2968 – 2978; d) J. A. Magee, T. Araki, S. Patil, T. Ehrig, L. True, P. A. Humphrey, W. J. Catalona, M. A. Watson, J. Milbrandt, *Cancer Res.* **2001**, *61*, 5692 – 5696; e) A. Tsuji, A. Torres-Rosado, T. Arai, M. M. Le Beau, R. S. Lemons, S. H. Chou, K. Kurachi, *J. Biol. Chem.* **1991**, *266*, 16948 – 16953.

[20] A. P. Silverman, M. S. Kariolis, J. R. Cochran, *J. Mol. Recognit.* **2011**, *24*, 127 – 135.

[21] A. Heitz, L. Chiche, D. Le-Nguyen, B. Castro, *Eur. J. Biochem.* **1995**, *233*, 837 – 846.

[22] A. Wentzel, A. Christmann, R. Kratzner, H. Kolmar, *J. Biol. Chem.* **1999**, *274*, 21037 – 21043.

[23] M. Empting, O. Avrutina, R. Meusinger, S. Fabritz, M. Reinwarth, M. Biesalski, S. Voigt, G. Buntkowsky, H. Kolmar, *Angew. Chem.* **2011**, *123*, 5313 – 5317; *Angew. Chem. Int. Ed.* **2011**, *50*, 5207 – 5211.

[24] K. F. Geoghegan, M. J. Emery, W. H. Martin, A. S. McColl, G. O. Daumy, *Bioconjugate Chem.* **1993**, *4*, 537 – 544.

[25] a) K. F. Geoghegan, J. G. Stroh, *Bioconjugate Chem.* **1992**, *3*, 138 – 146; b) G. Yu, A. R. Bayer, M. M. Galloway, K. J. Korshavn, C. G. Fry, F. N. Keutsch, *Environ. Sci. Technol.* **2011**, *45*, 6336 – 6342.

[26] A. P. Silverman, A. M. Levin, J. L. Lahti, J. R. Cochran, *J. Mol. Biol.* **2009**, *385*, 1064 – 1075.

[27] a) M. Cemazar, N. L. Daly, S. Haggblad, K. P. Lo, E. Yulyaningsih, D. J. Craik, *J. Biol. Chem.* **2006**, *281*, 8224 – 8232; b) R. Krätzner, J. E. Debreczeni, T. Pape, T. R. Schneider, A. Wentzel, H. Kolmar, G. M. Sheldrick, I. Uson, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2005**, *61*, 1255 – 1262.

[28] T. E. Creighton, C. J. Bagley, L. Cooper, N. J. Darby, R. B. Freedman, J. Kemmink, A. Sheikh, *J. Mol. Biol.* **1993**, *232*, 1176 – 1196.

[29] a) J. P. Tam, C. R. Wu, W. Liu, J. W. Zhang, *J. Am. Chem. Soc.* **1991**, *113*, 6657 – 6662; b) T. J. Wallace, *J. Am. Chem. Soc.* **1964**, *86*, 2018 – 2021; c) Z. Wu, D. M. Hoover, D. Yang, C. Boulegue, F. Santamaria, J. J. Oppenheim, J. Lubkowski, W. Lu, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 8880 – 8885.

[30] a) D. E. Callihan, T. M. Logan, *J. Mol. Biol.* **1999**, *285*, 2161 – 2175; b) J. Tam, X. Dong, C. Wu, *Lett. Pept. Sci.* **1999**, *6*, 265 – 273.

[31] a) D. Samuel, T. K. Kumar, G. Ganesh, G. Jayaraman, P. W. Yang, M. M. Chang, V. D. Trivedi, S. L. Wang, K. C. Hwang, D. K. Chang, C. Yu, *Protein Sci.* **2000**, *9*, 344 – 352; b) C. Lange, R. Rudolph, *Curr. Pharm. Biotechnol.* **2009**, *10*, 408 – 414.

[32] a) T. Shi, D. L. Rabenstein, *J. Am. Chem. Soc.* **2000**, *122*, 6809 – 6815; b) J. L. Huang, M. E. Noss, K. M. Schmidt, L. Murray, M. R. Bunagan, *Chem. Commun.* **2011**, *47*, 8007 – 8009; c) A. A. Miloslavina, E. Leipold, M. Kijas, A. Stark, S. H. Heinemann, D. Imhof, *J. Pept. Sci.* **2009**, *15*, 72 – 77.

[33] M. Erlandsson, M. Hällbrink, *Int. J. Pept. Res. Ther.* **2005**, *11*, 261 – 265.

[34] a) P. Lambert, H. Kuroda, N. Chino, T. X. Watanabe, T. Kimura, S. Sakakibara, *Biochem. Biophys. Res. Commun.* **1990**, *170*, 684 – 690; b) C. Miller, E. Moczydlowski, R. Latorre, M. Phillips, *Nature* **1985**, *313*, 316 – 318.

[35] a) F. Aslund, K. D. Berndt, A. Holmgren, *J. Biol. Chem.* **1997**, *272*, 30780 – 30786; b) P. M. Wood, *Biochem. J.* **1988**, *253*, 287 – 289.

[36] a) M. Mammen, S.-K. Choi, G. M. Whitesides, *Angew. Chem.* **1998**, *110*, 2908 – 2953; *Angew. Chem. Int. Ed.* **1998**, *37*, 2754 – 2794; b) V. Martos, P.

Castreno, J. Valero, J. de Mendoza, *Curr. Opin. Chem. Biol.* **2008**, *12*, 698 – 706; c) F. Song, W. C. Chan, *Nanotechnology* **2011**, *22*, 494006.
[37] a) N. J. Agard, J. A. Prescher, C. R. Bertozzi, *J. Am. Chem. Soc.* **2004**, *126*, 15046 – 15047; b) S. Fabritz, D. Heyl, V. Bagutski, M. Empting, E. Rikowski, H. Frauendorf, I. Balog, W. D. Fessner, J. J. Schneider, O. Avrutina, H. Kolmar, *Org. Biomol. Chem.* **2010**, *8*, 2212 – 2218; c) M. Galibert, O. Renaudet, P. Dumy, D. Boturyn, *Angew. Chem.* **2011**, *123*, 1941 – 1944; *Angew. Chem. Int. Ed.* **2011**, *50*, 1901 – 1904; d) T. S. Zatsepin, D. A. Stetsenko, A. A. Arzumanov, E. A. Romanova, M. J. Gait, T. S. Oretskaya, *Bioconjugate Chem.* **2002**, *13*, 822 – 830; e) S. Fabritz, S. Hörner, D. Könning, M. Empting, M. Reinwarth, C. Dietz, B. Glotzbach, H. Frauendorf, H. Kolmar, O. Avrutina, *Org. Biomol. Chem.* **2012**, *10*, 6287 – 6293.
[38] a) I. I. Geschwind, C. Hao Li, *Biochim. Biophys. Acta* **1954**, *15*, 442 – 443; b) J. Brask, K. J. Jensen, *J. Pept. Sci.* **2000**, *6*, 290 – 299.
[39] a) Y. Shechter, Y. Burstein, A. Patchornik, *Biochemistry* **1975**, *14*, 4497 – 4503; b) R. B. Yamasaki, D. T. Osuga, R. E. Feeney, *Anal. Biochem.* **1982**, *126*, 183 – 189.
[40] I. S. Bengelsdorf, *J. Am. Chem. Soc.* **1953**, *75*, 3138 – 3140.
[41] R. G. Boy, W. Mier, E. M. Nothelfer, A. Altmann, M. Eisenhut, H. Kolmar, M. Tomaszowski, S. Kramer, U. Haberkorn, *Mol. Imaging Biol.* **2010**, *12*, 377 – 385.
[42] J. F. Morrison, *Biochim. Biophys. Acta Enzymol.* **1969**, *185*, 269 – 286.

Received: September 19, 2012
Published online on December 11, 2012

3.2 Combinatorial Optimization of Cystine-knot Peptides towards High-Affininty Inhibitors of Human Matriptase-1

Title

Combinatorial Optimization of Cystine-Knot Peptides towards High-Affinity Inhibitors of Human Matriptase-1

Authors

Bernhard Glotzbach ,Michael Reinwarth, Niklas Weber, Sebastian Fabritz, Michael Tomaszowski, Heiko Fittler, Andreas Christmann, Olga Avrutina, Harald Kolmar

Bibliographical Data

PLoS ONE, **2013**; 8(10): e76956

DOI: 10.1371/journal.pone.0076956

Received: 13 June 2013 / Accepted: 27 August 2013 / Published: 11 October 2013

Contributions by Michael Reinwarth

- Synthesis of all miniproteins in multi-milligram scale
- Acquisition of analytical data of cystine-knot miniproteins
- Design and interpretation of cellbiological experiments
- Creation of manuscript and several figures

OPEN ACCESS Freely available online

PLOS ONE

Combinatorial Optimization of Cystine-Knot Peptides towards High-Affinity Inhibitors of Human Matriptase-1

Bernhard Glotzbach[1☯], Michael Reinwarth[1☯], Niklas Weber[1], Sebastian Fabritz[2], Michael Tomaszowski[1], Heiko Fittler[1], Andreas Christmann[1], Olga Avrutina[1], Harald Kolmar[1]*

1 Clemens-Schöpf-Institut für Organische Chemie und Biochemie, Technische Universität Darmstadt, Darmstadt, Germany, **2** AB SCIEX Germany GmbH, Darmstadt, Germany

Abstract

Cystine-knot miniproteins define a class of bioactive molecules with several thousand natural members. Their eponymous motif comprises a rigid structured core formed by six disulfide-connected cysteine residues, which accounts for its exceptional stability towards thermic or proteolytic degradation. Since they display a remarkable sequence tolerance within their disulfide-connected loops, these molecules are considered promising frameworks for peptide-based pharmaceuticals. Natural open-chain cystine-knot trypsin inhibitors of the MCoTI (*Momordica cochinchinensis* trypsin inhibitor) and SOTI (*Spinacia oleracea* trypsin inhibitor) families served as starting points for the generation of inhibitors of matriptase-1, a type II transmembrane serine protease with possible clinical relevance in cancer and arthritic therapy. Yeast surface-displayed libraries of miniproteins were used to select unique and potent matriptase-1 inhibitors. To this end, a knowledge-based library design was applied that makes use of detailed information on binding and folding behavior of cystine-knot peptides. Five inhibitor variants, four of the MCoTI family and one of the SOTI family, were identified, chemically synthesized and oxidatively folded towards the bioactive conformation. Enzyme assays revealed inhibition constants in the low nanomolar range for all candidates. One subnanomolar binder ($K_i = 0.83$ nM) with an inverted selectivity towards trypsin and matriptase-1 was identified.

Citation: Glotzbach B, Reinwarth M, Weber N, Fabritz S, Tomaszowski M, et al. (2013) Combinatorial Optimization of Cystine-Knot Peptides towards High-Affinity Inhibitors of Human Matriptase-1. PLoS ONE 8(10): e76956. doi:10.1371/journal.pone.0076956

Editor: Jose M. Sanchez-Ruiz, Universidad de Granada, Spain

Received June 13, 2013; **Accepted** August 27, 2013; **Published** October 11, 2013

Funding: This work was funded by Deutsche Forschungsgemeinschaft (http://www.dfg.de/en/index.jsp) through grants KO-1390/9-1 and SPP 1623 grant 1390/10-1. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors thank AB SCIEX Germany GmbH, Darmstadt for providing equipment and expertise in MS measurements of the cystine-knot peptides. Author Sebastian Fabritz is an employee of AB SCIEX Germany GmbH. There are no patents, products in development or marketed products to declare. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: Kolmar@Biochemie-TUD.de

☯ These authors contributed equally to this work.

Introduction

Cystine-knot peptides, often referred to as knottins, can be considered as one of Nature's combinatorial libraries [1–4]. These peptides have been identified in various organisms, among them fungi, plantae, porifera, mollusca, arthropoda, and vertebrata. While they share a common fold, they display a notably large diversity within the primary structure of flanking loops that is also correlated with a diversity of biological activities [2–5]. Their amide backbone of about 30 to 40 amino acid residues is compacted by three disulfide bonds which form the characteristic mechanically interlocked structure [6]. Three β-strands linked through three disulfide bonds define their structural core, where the ring-forming connection of CysI to CysIV and CysII to CysV is penetrated by a third cystine between CysIII and CysVI (Figure 1) [1–4]. NMR measurements of dynamics of backbone NH groups revealed high structural rigidity [7]. Considering the extensive network of hydrogen bonds which permeates the inner core, especially *via* the *β*-strands, thus adding a substantial thermodynamic stability, the cystine-knot motif displays an exceptional structural and thermal robustness [8–10].

Trypsin inhibitors isolated from the bitter gourd *Momordica cochinchinensis* (MCoTI, Figure 1A) and the squirting cucumber *Ecballium elaterium* (EETI) are prominent members of the ICK (inhibitor cystine-knot) family. Both share the typical architecture of an ICK peptide with the functional loop comprising six amino acids located between CysI and CysII (Figure 1) [3,11]. In contrast, recently reported miniproteins isolated from spinach *Spinacia oleracea* (SOTI I–III, Figure 1B) have shown no similarity to known plant protease inhibitors, but to antimicrobial peptides from the seeds of *Mirabilis jalapa* with the inhibitory loop located between CysV and CysVI (Figure 1) [12,13]. Structural information is available for the members of both inhibitor families [13–17].

Sequence and structure alignments of members of a respective miniprotein family reveal a conserved structural core, while the surface-exposed loops possess a high flexibility in terms of primary structure [3]. Thus, through substitution of surface-exposed residues bioactive variants can be generated that can serve as tailor-made compounds for potential diagnostic and therapeutic applications [10,18–20]. Several knottins have already been optimized by rational design or combinatorial library screening towards binding to targets of medical relevance [18,21–32]. For example, a MCoTI-II-derived miniprotein comprising a non-native hydrazone macrocyclization motif was reported to simul-

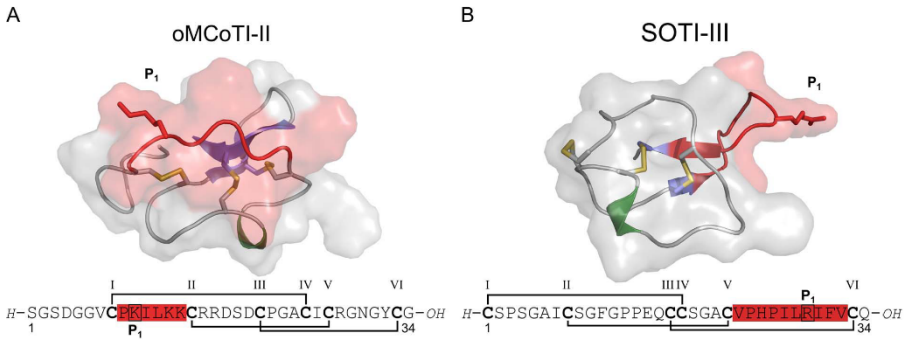


Figure 1. Sequences and structures of cystine-knot trypsin inhibitors. (**A**) Knottin oMCoTI-II (pdb: 1ha9). (**B**) SOTI-III (pdb: 4aor). Secondary structure is shown as cartoon with surface, and cysteine residues are depicted as yellow sticks; protease-binding regions (or inhibitor loops) are depicted in red, β-sheets - in blue, and α-helices - in green. Cystine-forming residues are marked bold, and the numbering of respective cysteines is according to their appearance in the sequence.
doi:10.1371/journal.pone.0076956.g001

taneously inhibit all four monomers of human mast cell tryptase β, a protease of clinical relevance related to allergic asthma [27,28]. Several rounds of directed evolution and rational design of the scorpion-derived miniprotein Leiurotoxin I from *Leiurus quinquestriatus hebraeus* resulted in its enhanced binding to gp120 of the viral particle of HIV, thus inhibiting cell entry [25,26,33]. Furthermore, cancer-related integrins have been successfully labeled *in vivo* with radioactive ^{64}Cu and ^{111}In *via* selective targeting with knottins containing an integrin-binding RGD motif and used for PET (positron emission tomography) and SPECT (single-photon emission computed tomography) imaging [21–24].

Knottins are readily accessible both by recombinant production and SPPS (solid-phase peptide synthesis) [3]. Indeed, obvious difficulties arising upon on-support chain assembly can be easily overcome using the wide-ranging repertoire of modern peptide synthesis, and the crucial step, regioselective formation of a tridisulfide pattern, can be efficiently controlled using optimized oxidation conditions [3,34].

Matriptase-1, a TTSP (type II transmembrane serine protease) of about 855 amino acids, belongs to the family of S1 trypsin-like proteases [35,36]. It combines an amino terminal hydrophobic transmembrane region with an extracellular section of several domains, among them a trypsin-like catalytic and a low-density lipoprotein region [35–37]. Autocatalytic activation of the zymogen is assisted by its cognate inhibitor HAI-1 (hepatocyte growth factor activator inhibitor-1) and does not depend on other proteases. To date, the mechanism of autocatalytic activation has not been fully understood [35,37–39]. Interestingly, matriptase-1 is also activated *via* acidification of the enzyme, therefore indicating its role in cellular acidosis [40]. Studies on knock-out mice have shown that matriptase-1 is essential for epidermal barrier functions, hence postnatal survival, as well as growth of hair follicles, and thymic homeostasis [41]. Moreover, matriptase-1 has been reported to be expressed not only in epithelial cells, but also in mast cells, B-cells, and blood monocytes [42–44]. Among its numerous substrates of which most are important for cell adhesion and tissue remodeling, processing of pro-uPA (pro-urokinase plasminogen activator) and pro-HGF (pro-hepatocyte growth factor) have been shown to be significantly involved in tumor growth and metastasis [45]. Expression rates of matriptase-1 were reported to reflect the degree of tumor progression in several types of cancerous cells, thus indicating a crucial role of this protease in tumor metastasis [46–48]. This was evidenced through various experiments, both *in vitro* and *in vivo*, in which the enzyme was inhibited [35,49–51]. The ratio of matriptase-1 and HAI-1, which is shifted towards matriptase-1 in cancer cells, is of major importance for tumor invasiveness [45,52,53]. Moreover, matriptase-1 has been reported to be implicated in a number of other diseases, among them osteoarthritis and atherosclerosis, and to induce cancer itself [42,54,55]. In conclusion, matriptase-1 has become a promising target for drug development. To date, only one peptide-based inhibitor of matriptase-1 with a picomolar K_i has been reported [56,57]. Despite its excellent inhibition constants against matriptase-1, this four-amino-acid peptide with the sequence *H*-R-Q-A-R-*Bt* (Bt stands for carboxy terminal benzothiazole substituent) displays a low selectivity. Since for *in vivo* experiments a high selectivity and serum half-life are indispensable, this inhibitor presumably is not suitable for experiments towards tumor targeting *in vivo*. Here we describe the isolation of selective cystine-knot peptides of high affinity from knowledge-based combinatorial miniprotein libraries and their functional characterization *in vitro* and in cell culture.

Materials and Methods

Media and Reagents

All media were prepared as previously reported [18,58,59]. YPD medium contained 20 g/L peptone, 20 g/L dextrose, and 10 g/L yeast extract. Selective SD-CAA medium incorporated 6.7 g/L yeast nitrogen base without amino acids, 20 g/L dextrose, 8.6 g/L $NaH_2PO_4 \cdot H_2O$, 5.4 g/L Na_2HPO_4, and 5 g/L Bacto casamino acids. SG-CAA medium was prepared similarly except for the addition of 100 mL/L polyethylene glycol 8000 (PEG 8000) and the substitution of dextrose by galactose. DYT medium contained 10 g/L yeast extract, 16 g/L tryptone, 5 g/L and 100 mg/L ampicillin. Phosphate-buffered saline (PBS) was composed of 8.1 g/L NaCl, 0.75 g/L KCl, 1.13 g/L Na_2HPO_4, and 0.27 g/L KH_2PO_4 at pH 7.4.

RPMI cell culture media (with and without phenol red) was supplemented with 10% (v/v) fetal calf serum (FCS) and antibiotics. These materials were purchased from Sigma-Aldrich.

Human matriptase-1 was produced recombinantly, autocatalytically activated and purified as previously reported [35,49,51,60]. Bovine pancreatic trypsin, thrombin and uPA were purchased from Sigma-Aldrich and Hepsin from R&D Systems.

Variant Cloning and Library Synthesis

For the initial display experiments of SOTI-III *wild type* and the yeast libraries based on the MCoTI-II and SOTI-III scaffold the encoding gene fragments were amplified by PCR with Taq polymerase with the use of primers with 50-bp overlap to the pCT plasmid up- or downstream of the *Nhe*I and *Bam*HI restriction sites, respectively. Positions for randomization in case of the SOTI-III library contained the NNK degenerate codon. For the MCoTI-II library, weighted randomization of respective residues was achieved upon synthesis using pre-made codon mixtures as described [61]. Amplified PCR products were purified by phenol/chloroform extraction. The vector was digested with *Nhe*I and *Bam*HI and purified *via* sucrose density gradient for homologous recombination in yeast. For the electroporation reaction 1–4 µg of linearized plasmid and 10–12 µg of insert were used [58]. After 1 h incubation (YPD medium, 30°C) library size was estimated by dilution plating. The yeast cells were transferred into selective SD-CAA medium, grown at 30°C to $OD_{600} = 10–12$ and split into new SD-CAA medium. Library stocks were stored at –80°C [58]. Yeast cells were induced in SG-CAA medium (starting OD_{600} of 0.1–0.2, 20°C, 48 h, 220 rpm).

Surface Binding Assays and Library Screening

Surface presentation of miniproteins was monitored by flow cytometry. $1 \cdot 10^7$ cells were labeled consecutively with 1:20 dilutions of anti-cMyc antibody (monoclonal, mouse, Abcam), anti-mouse IgG biotin conjugate (polyclonal, goat, Sigma-Aldrich), and Streptavidin, R-phycoerythrin conjugate (SPE) for 10 min on ice.

Protease binding assays and one-dimensional screenings of recombinant knottin libraries were conducted by incubation of knottin-presenting yeast cells with the respective biotinylated protease for 30 minutes on ice. Subsequently, the cells were resuspended in a 1:20 dilution of SPE for 10 min. The cells were analyzed in an Accuri C6 (Becton Dickinson) or were sorted using a MoFlo cell sorter. Sorting parameters were: trigger side scatter 650, PMT FL2 600, ex. 488 nm filter FL2 570/40. FCS files were analyzed using CFlow software or Summit 4.3, respectively.

For two-dimensional screening the yeast cells were consecutively incubated for 30 min at 0°C with 1:20 dilutions of each anti-cMyc antibody containing the desired concentration of biotinylated protease as well as a mixture of SPE and anti-mouse-IgG FITC (parameters: trigger side scatter 650, FL1 600, FL2 600).

Approximately 2×10^8 yeast cells were run through the flow cytometer at the first round of sorting. The selected cells were cultured after each screening round in SD-CAA medium. Next screening rounds were performed with at least 10 times the number of yeast cells collected in the previous round to ensure library diversity. Sort stringency was increased by reducing the protease concentration in subsequent screening rounds.

Plasmid DNA from positive clones was isolated and transformed into DH5α competent *E. coli* cells for plasmid amplification. DNA sequencing was performed using the oligonucleotide pCT-seq-lo.

Cell Inhibition Assay

Human prostate cancer cells (PC-3, Merck KGaA) were cultured in DMEM medium with 10% FCS at 37°C and 5% CO_2, washed with cation-free PBS and harvested by scraping. 1×10^5 cells were incubated in presence of 250 µM Bz-β-Ala-Gly-Arg-*p*NA·AcOH (American Diagnostica), which is a specific inhibitor of urokinase, and the inhibitor of interest in defined dilutions overnight. Product formation was monitored at 405 nm before and after incubation in a microplate reader. IC_{50} was calculated by non-linear regression using SigmaPlot 11.

Synthesis of Cystine-knot Miniproteins

Peptides were assembled using standard Fmoc-SPPS chemistry on a fully automated microwave-assisted CEM *Liberty*® peptide synthesizer. Peptide acids were generated using an Fmoc-Gln-preloaded TentaGel resin, whereas peptide amides were synthesized on a *ChemMatrix* Fmoc-Rink amide resin. After cleavage from the solid support, oxidative folding was conducted as recently reported [34]. About 40 mg of the corresponding lyophilized crude peptide were suspended in 500 µL acetonitrile and treated in an ultra-sonic bath for 5 min. Afterwards, 3500 µL of the folding mixture consisting of 10% (*v/v*) DMSO, 10% (*v/v*) TFE and guanidinium hydrochloride (GuHCl) (1 M) in aqueous sodium phosphate buffer (50 mM, pH 7) were added [34]. Reaction progress was monitored *via* analytical HPLC and ESI-MS (Figures S7 and S8) [34]. For termination of the reaction and purification of the bioactive miniprotein, the mixture was directly injected into a semi-preparative HPLC system.

RP-HPLC, LC-ESI-MS, and CD Spectroscopy

Analytical RP-HPLC was performed using a Varian LC 920 system equipped with a *Phenomenex* Synergi 4 µ Hydro-RP 80 Å (250×4.6 mm, 4 µm) column applying linear gradients of acetonitrile at a flow rate of 1 mL/min. Semi-preparative RP-HPLC purifications were performed using a Varian LC 940 system equipped with an axia-packed *Phenomenex* Luna C18 (250×21.2 mm, 5 µm, 100 Å) column applying linear acetonitrile gradients at a flow rate of 18 mL/min. Isocratic elution (10% eluent B over 2 (on analytical scale) or 5 min (on semi-preparative scale)) was followed by a linear gradient of 10→60% B (for MCoTI variants) or 10→80% B (for SOTI variants) over 20 min, respectively.

LC-MS was performed with a *Shimadzu* LC-MS 2020 equipped with a *Phenomenex* Jupiter C4 (50×1 mm, 5 µm, 300 Å) column using linear acetonitrile gradients at a flow rate of 0.2 mL/min (Figures S7 and S8). Isocratic elution (2% eluent B over 2 min) was followed by a linear gradient of 2→100% B over 10 min. Cystine-knot disulfide bond topology of MCoTI Var. 4 was confirmed using MS^3-technology (AB Sciex, 4000 QTRAP® LC/MS/MS System; data not shown).

CD spectroscopy was performed as previously reported [34]. The peptides were dissolved in 2 mM aqueous Na_2HPO_4 (pH 7) to a final concentration of 50 µM. The resulting spectra (Figure S9 and S10) were obtained through accumulation of 10 spectra each, using a 0.1 mm quartz cuvette at 0.5 nm steps.

Inhibition Assays

Protease inhibition assays which resulted in substrate-independent inhibition constants were performed as previously described [11,13,34,62].

Measurements were carried out in triplicate using a *Tecan Genios* ELISA reader. The normalized residual proteolytic activity (v/v_0) of proteases was determined using substrates Boc-QAR-*p*NA (250 µM, (pNA stands for *para*-nitro aniline), Boc-QAR-AMC (250 µM, AMC stands for amino-methyl coumarin) or Spectrozym tPA (250 µM, American Diagnostica, CH_3SO_2-D-CHT-Gly-Arg-pNA AcOH). Product formation was monitored after preincubation (30 min, RT) with inhibitor at different concentrations over 30 min by measuring the absorbance at 405 nm for

*p*NA substrates or the fluorescence emission for AMC substrates (ex. 360 nm, em. 465 nm), respectively. Selectivity data were carried out in duplicates with final protease concentrations of uPA and thrombin of 5 nM. In case of hepsin 50 mM Tris/HCl pH 9.0 was used as assay buffer.

Apparent inhibition constants (K_i^{app}) were calculated by fitting the Morrison equation (1) for tight-binding inhibitors to the relative reaction velocity using non-linear regression (Marquardt-Levenberg algorithm, Sigma Plot 11) [63].

$$\frac{v}{v_0} = 1 - \frac{\left(E_0 + I_0 + K_i^{app}\right) - \sqrt{\left(E_0 + I_0 + K_0\right) - 4E_0 I_0}}{2E_0} \quad (1)$$

$$K_i = \frac{K_i^{app}}{\left(1 + \frac{[S]}{K_M}\right)} \quad (2)$$

Substrate-independent inhibition constants K_i were calculated from K_i^{app} and K_m of the enzyme according to (2). The Michaelis-Menten constant K_m for the substrates and proteases were determined previously [49,64].

Results

Selection of Knottin Scaffolds

Since the overall structure of matriptase-1 is similar to trypsin, the preference for cleavage at basic residues at the P_1 position is maintained [49]. Hence, we considered trypsin-inhibiting miniproteins as a starting point for functional combinatory library design to isolate inhibitors of matriptase-1 [17]. From the plethora of miniproteins that are characterized to date, scaffolds were selected matching the following criteria: inhibitor of a trypsin-like protease, known three-dimensional structure, tolerance to variation of loop lengths and sequence, known mechanism of folding and disulfide bond formation, as well as availability through chemical and recombinant routes of synthesis [2–4]. Two different scaffold proteins have been selected based on the aforementioned requirements. The first selected scaffold was based on the spinach-derived inhibitor SOTI-III. The structure of this protease inhibitor has been recently elucidated by X-ray crystallography [13]. Since the inhibitor loop of SOTI-III is located between CysV and CysVI, this miniprotein is structurally and sequentially very distinct to MCoTI-II, which was chosen as second scaffold (Figure S1). This scaffold is based on miniproteins from the seeds of the squash plant *M. cochinchinensis*. This plant produces a number of miniprotein-based trypsin inhibitors, both backbone-cyclized macrolactams and variants lacking this motif (so-called 'open-chain' variants), which are slightly different in their sequences [1,17]. To evaluate which of the natural MCoTI variants could serve as a scaffold for the generation of matriptase-1 inhibitors, natural inhibitors were isolated form the *M. cochinchinensis* seeds using known extraction procedures followed by HPLC separation (Figure S2) [17,65]. Miniproteins from various fractions were identified by ESI-MS and examined for inhibition of matriptase-1 (Table S1). MCoTI-II, a macrolactam-cyclized miniprotein consisting only of natural amino acids, was found to be the most efficient natural inhibitor of matriptase-1 and therefore chosen as starting scaffold. Synthetic open-chain MCoTI-II (oMCoTI) displayed a K_i^{app} similar to that of its cyclic counterpart (Table S1). Interestingly, SOTI-III is a less potent inhibitor of trypsin and did not display measurable inhibitory activity against matriptase-1 (Table 1).

Table 1. Inhibition constants of inhibitors studied in this work.

Inhibitor	K_i (Trypsin) (nM)	K_i (Matriptase-1) (nM)
SOTI-III *wt*	60.6±8.4	>1000
SOTI Var. 1	>1000	28.9±3.5
MCoTI-II *wt*	2.37±0.96	80.7±10.0
MCoTI Var. 1	31.7±4.3	4.4±0.6
MCoTI Var. 2	19.2±2.8	3.3±0.4
MCoTI Var. 3	22.3±3.0	7.8±1.0
MCoTI Var. 4	35.8±4.7	0.83±0.14
S1[a] [51]	118±16	7.1±0.87
S2[a] [51]	544±74	28.2±3.5

[a]Structural information for reference compounds S1 and S2 are depicted in Figure S3.

doi:10.1371/journal.pone.0076956.t001

SOTI-III-based Library Screening

To obtain knottin-based matriptase-1 binders, yeast surface display was chosen as its applicability to the screening of cystine-knot-based peptide libraries has been already demonstrated [22,58,59,66]. To this end, the SOTI-III *wild type* or library-encoding DNA was genetically fused to the *Saccharomyces cerevisiae* Aga2p coding sequence. The resulting constructs are under control of the galactose promoter [22]. Induction with galactose yields a fusion protein consisting of Aga2p, a glycine-serine linker, an HA-epitope, the miniprotein, and a cMyc epitope (Figure 2A) [58,66–68]. The fusion is covalently bound to the surface-anchored Aga1p [58,66]. Functional display of SOTI-III *wt* was shown by binding of biotinylated bovine pancreatic trypsin followed by flow cytometric analysis (Figure 2B). After verification of functional display of the *wild type* miniprotein, the inhibitor loop was randomized by PCR using oligonucleotides with NNK codon randomization (materials and methods section). All ten loop residues of SOTI-III were considered for full randomization including the P_1 residue arginine, since for optimized matriptase-1 binding the P_1 residue may be shifted to another position within the inhibitor loop. The resulting miniprotein library had a clonal diversity of 2×10^8.

To isolate matriptase-1-binding SOTI-III variants, four consecutive fluorescence-activated cell sorting (FACS) screening rounds were performed (Figure 2, Figure 2C for flow cytometry screening, and materials and methods section) and one dominant clone was isolated after four sorting rounds (Figure 2D).

Subsequently, an alanine scan was conducted to determine the essential arginine residues within the inhibitor loop that contained four arginine residues. As a result, Arg29 and Arg32 were found imperative for binding and bioactivity, while Arg30 and Arg35 were dispensable without major loss of binding (Figure S4). The SOTI-based cystine-knot peptide Var. 1 was synthesized chemically using microwave-assisted Fmoc-SPPS followed by oxidative folding and HPLC purification yielding the bioactive peptide (Table S2). Subsequent inhibition assays revealed the inhibition constants against matriptase-1 and trypsin as 28.9 nM and >1 μM, respectively (Table 1).

MCoTI-II-based Library Design

Encouraged by these promising results, we further optimized the library design towards more potent cystine-knot inhibitors of matriptase-1. A codon-based randomization of the oMCoTI-II

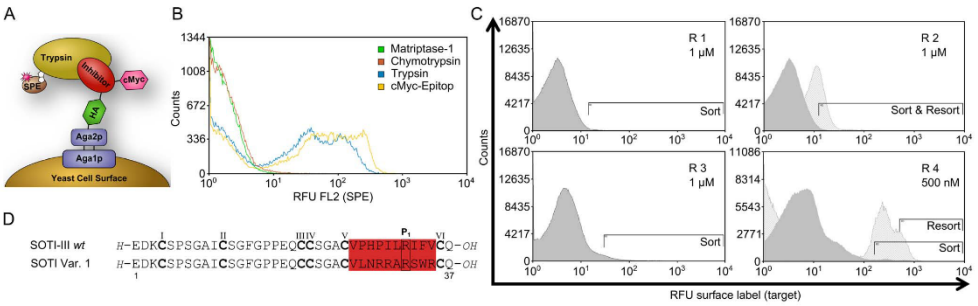


Figure 2. Yeast surface display of SOTI-III *wild type* and screening against matriptase-1. (**A**) Schematic illustration of Aga1p/Aga2p surface-displayed inhibitor (red) flanked by the amino terminal HA (Human influenza hemagglutinin) epitope (green) and the carboxy terminal cMyc epitope (purple). Functional display of the inhibitor is monitored by incubation with biotinylated trypsin followed by fluorescence labeling with streptavidin, R-phycoreythrin conjugate (SPE). (**B**) FACS histogram overlay of yeast surface presented SOTI-III *wild type* labeled with anti-cMyc antibody (yellow), trypsin (blue), matriptase-1 (green) and chymotrypsin (brown). (**C**) FACS overlays of matriptase-1 binder enrichment. The sorting round (R) and the matriptase-1 concentration used in each round (µM) is given in the figures. Dark grey: FACS histogram during sorting. Light grey: FACS histogram during resort (only rounds 2 and 4). (**D**) Sequence alignment of SOTI-III *wild type* and matriptase-1-binding SOTI variant 1. Randomized residues are colored in red. Cysteines are depicted in bold letters, while cystine connections are omitted for clarity.
doi:10.1371/journal.pone.0076956.g002

scaffold was used for library generation (Figure 3A and 3B), which included the inhibitor loop and neighboring residues that may contribute to target binding [61]. It is well known that a proline is required at position P_2 (amino-terminal to P_1, Figure 1) of the inhibitor loop [2,4]. Thus, Pro5 was not modified since it is essential for the formation of the six-residue canonical inhibitor loop conformation that is found in many protease inhibitors [1,2,4]. Codon 6 was randomized to code for Arg or Lys (50% each), and positions 7–10 were randomized to code for the full set of 19 canonical amino acids, excluding cysteine, using a codon-based randomization scheme (Figure 3A) [22]. In addition, neighboring residues were also included into the variegation scheme to enable improved subsite binding that may contribute to both enhanced affinity and specificity. Since these residues outside the inhibitor loop may be of relevance for oMCoTI-II folding and stability, simultaneous full randomization was avoided by maintaining the original residue at each position for 50% of the variants. As a consequence, in approximately 3% of the variants all five original amino acids that are located adjacent to the inhibitor loop are expected to be preserved and the average number of residue replacements was expected to be 7 (Figure S5). In oMCoTI-II, the carboxy-terminal loop is located adjacent to the inhibitor loop and therefore can affect target binding. Tolerance of this loop region towards amino acid exchanges has been extensively investigated for the structurally similar knottin EETI [18,29,69,70]. This loop region is thought to be involved in the early folding process of the miniprotein via formation of a type II β-turn [2,4,11,18,70]. Since this loop sequence is a folding determinant, only moderate sequence variations were included by randomizing each position to 10%. Thus, over 50% of the variants can be expected to have none or one amino acid exchange within that region (Figure S5). The same moderate mutagenesis scheme was applied for D14 and D16 that are conserved in the ICK family of miniproteins and are involved in stabilization of the oMCoTI scaffold [2,4]. As the active site of matriptase-1 is negatively charged, it may be beneficial for binding to allow replacement of these residues [49]. Overall, the randomization scheme applied here includes 17 out of 30 residues. However, on average only 6 to 8 of the 17 residues are expected to be changed in each variant and four of these are most likely located within the inhibitor loop.

MCoTI-II Library Screening

To evaluate the feasibility of library design that includes 17 of 30 residues in the randomization scheme, two relatively small yeast libraries with a diversity of 2×10^6 and 2×10^7 clones, respectively, were independently constructed from the same synthetic library DNA and screened separately. After two to four rounds of screening, matriptase-1-binding populations were enriched. Individual matriptase-1-binding clones were identified using flow cytometry (Figure 3C). DNA sequences were obtained (10 from the screen of the library with a diversity of 2×10^6 clones as well as 12 of the 3rd and 16 out of the 4th round of the library containing 2×10^7 clones, respectively; Figure S6). From these, four binders were selected for detailed investigations (Figure 3D) that were independently identified several times in screening rounds three and four or displayed high affinity binding upon yeast cell surface affinity titration (Figure S6).

To determine the inhibition constants, chemical synthesis and oxidative folding of the putatively inhibiting cystine-knot peptides were performed as previously reported (Table S2) [34]. Correct fold of the miniproteins was proven through bioactivity, since it is known that knottins of the ICK family displaying an incorrect disulfide connectivity show a decreased inhibitory efficiency [3,11,15,30,34]. Moreover, CD spectra of SOTI *wt*, SOTI Var. 1, MCoTI *wt* and MCoTI Var. 4 indicated β-sheet formation (Figures S9 and S10). Disulfide bond connectivities were confirmed by MS^3 mass spectrometry for MCoTI Var. 4 via continuous injection of a 3 µM solution of the miniprotein at a flow rate of 10 µL/min into an ABSCIEX 4000 QTRAP® LC/MS/MS system (data not shown) [71]. Inhibition constants in the low nanomolar to sub-nanomolar range were obtained for all MCoTI-based miniproteins (Table 1). An additionally performed selectivity study for the best MCoTI-based inhibitor candidate Var. 4 revealed inhibition constants K_i >10 µM against thrombin, uPA, and hepsin (Table 2). Moreover, inhibitory activity for matriptase-1 was approximately fortyfold higher than for trypsin (Table 1).

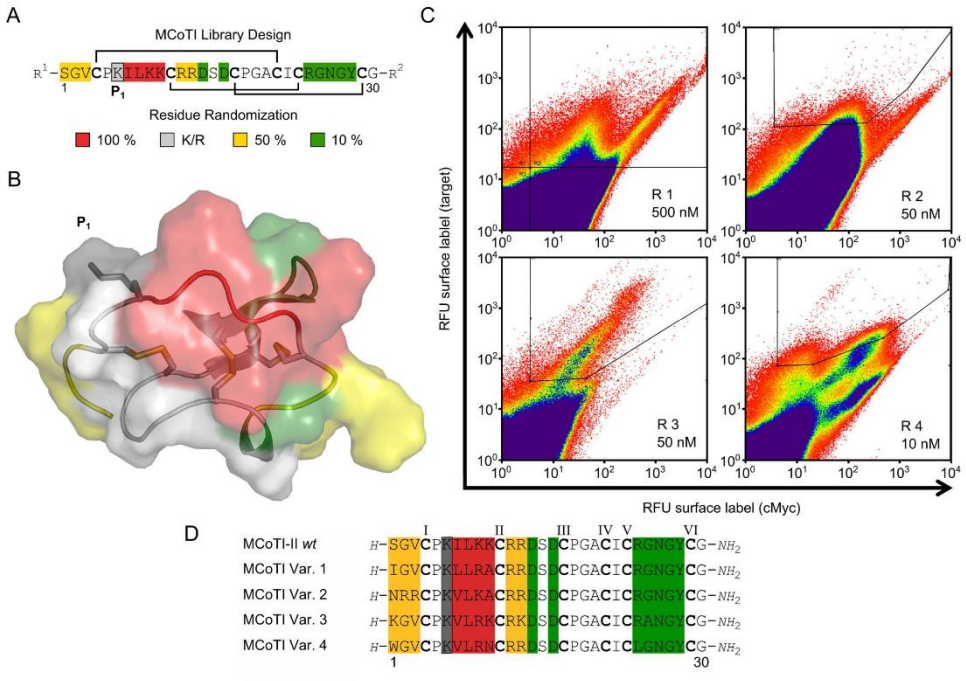


Figure 3. Summary of MCoTI-II-based library design and screening against matriptase-1. (**A**) Sequence of open-chain MCoTI-II *wild type*. Cysteines are depicted in bold letters. R^1 represents the amino-terminal flanking sequence, including the HA-epitope. R^2 represents the carboxy-terminal flanking sequence, including the cMyc-epitope. Codon randomization for (**A**), (**B**), and (**D**) as indicated by color (at pos. 6 only Lys or Arg was allowed, grey). (**B**) Secondary structure of MCoTI-II is shown as cartoon with surface, cysteine residues are shown as orange sticks. (**C**) FACS histograms showing four rounds of sorting with decreasing target concentration for enrichment of matriptase-1 binders. R denotes the sort round with the concentration of matriptase-1 indicated. Actual sort gates are shown. (**D**) Sequence alignment of matriptase-1-binding MCoTI variants. Cysteines are numbered according to the appearance in the sequence and depicted in bold letters, while cystine connections are omitted for clarity.
doi:10.1371/journal.pone.0076956.g003

Inhibition of uPA Activation

Urokinase-type plasminogen activator (uPA) causes the degradation of the extracellular matrix and plays a critical role in tumor invasion and metastasis [72,73]. It was shown that activation of receptor-bound pro-uPA is affected by matriptase-1, which results in a decreased ability of uPAexpressing tumor cells to invade an extracellular matrix layer upon inhibition of membrane-bound matriptase-1 [72]. To investigate the inhibitory activity of the newly isolated matriptase-1 inhibitors on pro-uPA activation, a dose-response assay of uPA activity was performed in cell culture with SOTI-based variant (Var. 1) and the most potent MCoTI-based inhibitor (Var. 4) on human prostate carcinoma cancer cells (PC-3), as a upregulation of matriptase-1 expression level has been reported for this cell line [45,69,72].

Table 2. Selectivity profile of MCoTI-based inhibitor Var. 4.

Protease	K_i (nM)
Trypsin	35.8±4.7
Matriptase-1	0.83±0.1
Thrombin	>10000[a]
Urokinase	>10000[a]
Hepsin	>10000[a]

[a]No inhibition was observed at 10 µM inhibitor concentration.
doi:10.1371/journal.pone.0076956.t002

For the indirect determination of the IC_{50} of SOTI Var. 1 and MCoTI Var. 4 on the surface of these cancer cells, the turnover of an uPA substrate was monitored. Pro-uPA is activated through non-inhibited matriptase-1 and substrate turnover was measured and compared to the previously reported small molecule inhibitor S1 of matriptase-1 (Figure 4) [51]. In this experimental setting, the MCoTI-based inhibitor Var. 4 (K_i = 0.83 nM) exhibited an IC_{50} of 213 nM, while SOTI-III derived inhibitor Var. 1 displayed only minor activity. S1 a small-molecule inhibitor (Figure S3) that has been identified recently as potent matriptase-1 inhibitor with an K_i in the single digit nanomolar range was used as reference compound that displayed an tenfold higher IC_{50} value than MCoTI-based inhibitor Var. 4 in this assay [51]. For control SOTI *wt* was also applied in this experimental setting at a

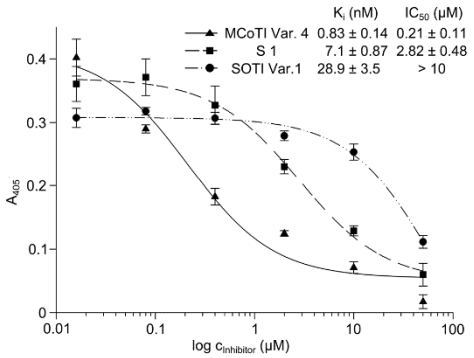


Figure 4. Inhibition assay of uPA activation by matriptase-1 on the surface of PC-3 cells. Depicted is the logarithmic inhibitor concentration against the absorption at 405 nm.
doi:10.1371/journal.pone.0076956.g004

concentration of 10 µM, displaying no inhibition of either matriptase-1 or uPA.

Discussion

For the isolation of miniprotein-based inhibitors by combinatorial library screening the design of the variant library is a crucial step. We chose a knowledge-based strategy that takes into account the expected contribution to target binding, as well as the natural variability and the contribution to structure and folding of each residue at each position. While we followed a classical variegation scheme for SOTI-III with a full randomization that is restricted to the carboxy terminal loop, a position-specific randomization scheme was applied for oMCoTI-II.

Screening of the SOTI-III library resulted in the isolation of a variant that displayed 29 nM K_i with respect to matriptase-1 inhibition and contained the sequence motif RRAR in the inhibitor loop. This agrees with the consensus sequence for matriptase-1 substrates and the highly potent inhibitor peptide *H*-R-Q-A-R-*Bt* [45,49,56,57]. Despite the fact that the absolute position of the P_1 arginine residue within the inhibitor loop remained unchanged, increased inhibitory activity towards matriptase-1 interestingly led to the total loss of trypsin inhibition. Hence, in comparison to the *wild type* miniprotein, the isolated SOTI-based matriptase-1 inhibitor (Var. 1) showed improved inhibitory activity and selectivity. Notably, the recently reported crystal structure of SOTI-III *wild type* revealed that out of the 10 carboxy terminal loop residues only 8 were in direct contact to trypsin [13]. Thus, exclusion of core-forming residues from the randomization scheme of SOTI-based inhibitor Var. 1 and generating a sub-library for the flanking residues might result in variants with further improved binding characteristics. Moreover, experiments on co-crystallization of matriptase-1 and SOTI Var. 1 are required to understand the mode of interaction of protease and inhibitor. Additionally, it would be beneficial to gain further knowledge on whether the conformational constraints that are imposed on the SOTI-III *wild type* inhibitor loop *via* integration into the cystine-knot scaffold are also conserved in the matriptase-1 inhibitor Var. 1. Assuming an unchanged mode of action, the knottin-based peptide acts as a matriptase-1 inhibitor rather than as a substrate that is readily and irreversibly cleaved.

Screening of the oMCoTI-scaffold-derived library resulted in several inhibitors that all displayed K_i values in low nanomolar to sub-nanomolar range. Despite the fact that the library diversity was 10-fold lower than for the SOTI scaffold and with 2×10^7 clones relatively small, more potent binders were isolated corroborating the concept of knowledge-based library design. It should also be noted that oMCoTI-II *wt* in contrast to SOTI-III *wt* already displayed inhibitory activity against matriptase-1 and therefore may be the more suitable scaffold for optimization towards matriptase-1-binding and inhibition. Branched aliphatic residues were observed at the $P_{1'}$ position of isolated matriptase-1-binding oMCoTI library variants, while leucine was the preferred $P_{2'}$-positioned amino acid. In contrast to reported substrates and inhibitors, lysine was obviously favored over an arginine residue at the P_1 position (Figure 3D) [45,56]. While amino acid residue 1 displayed a large variability, replacements of the 'GV-motif' at positions 2 and 3 rarely occurred, demonstrating the importance of these residues for binding and/or folding (Figure 3D). Whereas no substitutions within loop 2 (flanked by CysII and CysIII) were observed, Arg24 was exchanged for leucine in the most potent inhibitor Var. 4 (Figure 3D). It remains to be elucidated, whether this residue replacement contributes to enhanced inhibition.

To investigate the inhibition of pro-uPA activation cell culture upon matriptase-1 inhibition, miniproteins SOTI Var. 1 and MCoTI Var. 4 as well as reference compound S1 were applied to human pancreatic PC-3 cells (Figure 4) [45,72]. MCoTI-based knottin Var. 4 that had a subnanomolar K_i towards matriptase-1 also displayed the lowest IC_{50} with respect to the inhibition of proteolytic activity in a PC-3 cell line [74–76]. This indicates that inhibitor-mediated reduction of matriptase-1 activity contributes to the decrease of uPA activity. IC_{50} values ranged from a nanomolar to micromolar range and MCoTI-based inhibitor Var. 4 was found to be 10-fold more potent than recently described peptidomimetic small-molecule inhibitors (Figure 4) [51].

All three inhibitors investigated displayed IC_{50} values of protease inhibition on PC-3 cells more than 100-fold higher compared to their K_i of matriptase-1 inhibition. This discrepancy may arise from the complicated situation in cell culture since matriptase-1 activity is regulated by the cognate natural tight-binding inhibitor HAI-1. Co-expression of HAI-1 and matriptase-1 suppresses matriptase-1 proteolytic activity. Interestingly, HAI-1 has also been considered to be required for activation of matriptase-1 and to be involved in its expression and autoprocessing [38,39,41,57,77]. Moreover, absence of HAI-1 seems to cause rapid turnover of active matriptase-1 [57,77]. Hence, the complicated conditions in the cell-culture media, in the cell and on its surface may account for the observed differences of K_i and IC_{50}.

In recent years, matriptase-1 has attracted keen scientific interest as a target for the development of inhibitors. Steinmetzer and coworkers reported small molecule inhibitors that display similar potency and selectivity *in vitro* as well as in cell-based assays as the miniproteins generated in this study [51,76]. In addition, two types of peptidic matriptase-1 inhibitors have been identified to date [49,56]. The short substrate-derived inhibitor *H*-R-Q-A-R-*Bt* displays an inhibition constant in the double-digit picomolar range. Due to the small size and susceptibility to proteolytic degradation, *in vivo* half-life can be expected to be short. Moreover, the universal sequence is not selective for matriptase-1, but inhibits various proteases in the pico- to nanomolar range [56]. Compared to the tetrapeptide, the sunflower trypsin inhibitor (SFTI)-based matriptase-1 inhibitor that has been described recently has an increased size (14 residues) combined with a constrained structure, thus being potentially more stable and

applicable for *in vivo* experiments [49]. Recently, Daly and coworkers obtained by rational design and positional scanning mutagenesis a cyclic MCoTI-II variant with subnanomolar inhibition constant of matriptase-1 that was found to be more potent than SFTI derived variants, corroborating the notion that the McoTI-II miniprotein scaffold provides an excellent structural environment for the development of potent and selective matriptase inhibitors [78]. Serum stability and potential oral availability has been shown for several knottins and it will be interesting to see whether the MCoTI-derived inhibitors display similar stability in cell culture and serum while maintaining activity and selectivity.

Knottins have been introduced by Cochran and coworkers as a new class of agents for imaging of tumor marker expression in living systems [18,24,79]. For example, ^{64}Cu-DOTA-conjugated knottin peptides were stable in mouse serum, and *in vivo* metabolite analysis showed minimal degradation in blood or tumor rendering this type of stable peptides very promising candidates as clinical diagnostics for a variety of cancers [79]. The spectrum of tumor targeting knottins, which is currently restricted to cystine-knot peptides containing integrin binding RGD sequences in their binding loops, can be extended by matriptase-1 binders for imaging applications. The miniproteins described in this study selectively detect cell-surface-exposed and enzymatically active matriptase-1 on tumor cells that is not complexed with the natural inhibitor HAI-1. In contrast, with one notable exception most antibodies cannot distinguish between the active and inactive form of matriptase-1, due to their binding to accessible epitopes that are not linked to the active site or conformational changes upon activation. Recently, Craik and coworkers showed that an active-site-specific, recombinant human antibody for matriptase-1 can be used to visualize the tumorigenic epithelium using near-infrared and single-photon emission computed tomography imaging, corroborating the notion that the active form of matriptase-1 is a tumorigenic biomarker [81]. Since matriptase-1 provides the major contribution to tumor invasion and progression, knottins selectively addressing the active site of matriptase-1 may become valuable tools for tumor imaging, particularly for the prediction of tumor invasiveness.

Conclusions

To conclude, we have proven the applicability of a knowledge-based miniprotein library design to the development of potent inhibitors of human matriptase-1 using a codon-based, weighted, and selective randomization scheme. A set of cystine-knot miniprotein variants was generated that included a relatively large number of residues that may contribute to binding while the average number and position-specific frequency of amino acid replacements was carefully controlled. As a consequence, screening of a relatively small library revealed (sub−) nanomolar inhibitors. Bioactivity was confirmed in cell culture through a dose-response inhibition assay on the surface of human cancer cells. Taking into consideration the high affinity and selectivity combined with the high general thermodynamic stability of miniproteins, the variants described here may become promising tools for applications in cancer diagnostics. *In vivo* experiments towards tumor targeting with labeled synthetic miniproteins are currently in progress.

Supporting Information

Figure S1 Sequences and structure alignment of cystine-knot trypsin inhibitors. Secondary structure of oMCoTI-II (light brown, pdb: 1ha9, upper left) and SOTI-III (light blue, pdb: 4aor, upper right) is shown as cartoon and cysteine residues are depicted as yellow sticks; protease-binding regions are depicted in red. Cystine-forming residues are marked bold, and the numbering of respective cysteines is according to their appearance in the sequence.
(PNG)

Figure S2 HPLC trace of MCoTI-variants isolated from seeds of *Momordica cochinchinensis*. x marks an unidentified peak. [β-Asp]-MCoTI-II possesses a β-aspartyl residue at position 4. Cyclic miniproteins were isolated from 5 g of homogenized seeds. Extraction was performed using 20 mL aqueous sodium acetate (20 mM, pH 4.5) at ambient temperature for 16 h. The suspension was filtrated and proteins were denatured with 20 mL aqueous acetone (40%, v/v), while the miniproteins remained their native conformation. After removal of acetone under reduced pressure, the suspension was filtrated and the filtrate was purified by semi-preparative HPLC using an axia-packed Phenomenex Luna C18 (250×21.2 mm, 5 µm, 100 Å) column applying linear acetonitrile gradients at a flow rate of 10 mL/min. Isocratic elution (10% eluent B over 5 was followed by a linear gradient of 10→55% B over 30 min.
(TIF)

Figure S3 Small-molecule inhibitors of matriptase-1 that were used as reference compounds.
(PNG)

Figure S4 Matriptase-1 binding analysis of miniprotein variants SOTI Var. 1 and MCoTI Var. 4 via flow cytometry. (**A**) Sequence of the isolated matriptase-1 inhibitors with randomized residues depicted in the according color. (**B**) Overlay of FACS histograms after labeling of miniprotein-displaying yeast cells with 1 µM of biotinylated matriptase-1 followed by incubation with Streptavidin, R-phycoerythrin conjugate.
(PNG)

Figure S5 Knottin library design. Expected distribution of the appearance of amino acid exchanges in loop 1 (red), flanking regions of loop 1 (yellow), and loop 4 (green). The calculation was performed assuming a binominal distribution function.
(PNG)

Figure S6 Sequence alignments of MCoTI variants isolated from two screening cycles. Amino acids marked in red are identical to those of the MCoTI-*wt*; amino acids highlighted in red are conserved for all aligned sequences. The blue frames show the consensus of at least two amino acids. The consensus sequence (bottom line) was calculated with a threshold of 0.5. Consensus sequence: upper-case letters indicate sequential identity, lower-case letters illustrate consensus. MCoTI *wt* was taken as lead sequence for the alignment. Sequences that were selected for chemical peptide synthesis and further studies are marked on the right.
(PNG)

Figure S7 HPLC and MS analysis of folded miniprotein SOTI Var. 1. (**A**) HPLC trace (10 to 80% B over 20 min) at 220 nm. (**B**) ESI-MS of peptide-containing fraction.
(TIF)

Figure S8 HPLC and MS analysis of MCoTI Var. 1. (**A**) HPLC trace (10 to 60% B over 20 min) at 220 nm. (**B**) ESI-MS of peptide-containing fraction.
(TIF)

Figure S9 CD spectroscopy of the reduced (unfolded) and oxidized (folded) variants of SOTI *wt* and SOTI Var. 1. Smoothed with the 'smooth' function of Sigma Plot 11.
(TIF)

Figure S10 CD spectroscopy of the reduced (unfolded) and oxidized (folded) variants of MCoTI *wt* and MCoTI Var. 4. Smoothed with the 'smooth' function of Sigma Plot 11.
(TIF)

Table S1 Apparent inhibition constants towards matriptase-1 of the isolated cyclic MCoTI variants.
(PDF)

Table S2 Characterization of synthetic miniproteins.
(PDF)

References

1. Chiche L, Heitz A, Gelly JC, Gracy J, Chau PT, et al. (2004) Squash inhibitors: from structural motifs to macrocyclic knottins. Curr Protein Pept Sci 5: 341–349.
2. Gracy J, Le-Nguyen D, Gelly JC, Kaas Q, Heitz A, et al. (2008) KNOTTIN: the knottin or inhibitor cystine knot scaffold in 2007. Nucleic Acids Res 36: D314–319.
3. Reinwarth M, Nasu D, Kolmar H, Avrutina O (2012) Chemical Synthesis, Backbone Cyclization and Oxidative Folding of Cystine-knot Peptides - Promising Scaffolds for Applications in Drug Design. Molecules 17: 12533–12552.
4. Gelly JC, Gracy J, Kaas Q, Le-Nguyen D, Heitz A, et al. (2004) The KNOTTIN website and database: a new information system dedicated to the knottin scaffold. Nucleic Acids Res 32: D156–159.
5. Kolmar H (2009) Biological diversity and therapeutic potential of natural and engineered cystine knot miniproteins. Curr Opin Pharmacol 9: 608–614.
6. Craik DJ, Daly NL, Waine C (2001) The cystine knot motif in toxins and implications for drug design. Toxicon 39: 43–60.
7. Puttamadappa SS, Jagadish K, Shekhtman A, Camarero JA (2010) Backbone dynamics of cyclotide MCoTI-I free and complexed with trypsin. Angew Chem Int Ed Engl 49: 7030–7034.
8. Wang CK, Hu SH, Martin JL, Sjogren T, Hajdu J, et al. (2009) Combined X-ray and NMR analysis of the stability of the cyclotide cystine knot fold that underpins its insecticidal activity and potential use as a drug scaffold. J Biol Chem 284: 10672–10683.
9. Craik DJ, Cemazar M, Wang CK, Daly NL (2006) The cyclotide family of circular miniproteins: nature's combinatorial peptide template. Biopolymers 84: 250–266.
10. Heitz A, Avrutina O, Le-Nguyen D, Diederichsen U, Hernandez JF, et al. (2008) Knottin cyclization: impact on structure and dynamics. BMC Struct Biol 8: 54.
11. Avrutina O, Schmoldt HU, Gabrijelcic-Geiger D, Le Nguyen D, Sommerhoff CP, et al. (2005) Trypsin inhibition by macrocyclic and open-chain variants of the squash inhibitor MCoTI-II. Biol Chem 386: 1301–1306.
12. Kowalska J, Pszczola K, Wilimowska-Pelc A, Lorenc-Kubis I, Zuziak E, et al. (2007) Trypsin inhibitors from the garden four o'clock (Mirabilis jalapa) and spinach (Spinacia oleracea) seeds: isolation, characterization and chemical synthesis. Phytochemistry 68: 1487–1496.
13. Glotzbach B, Schmelz S, Reinwarth M, Christmann A, Heinz DW, et al. (2013) Structural characterization of Spinacia oleracea trypsin inhibitor III (SOTI-III). Acta Crystallogr D Biol Crystallogr 69: 114–120.
14. Heitz A, Chiche L, Le-Nguyen D, Castro B (1989) 1H 2D NMR and distance geometry study of the folding of Ecballium elaterium trypsin inhibitor, a member of the squash inhibitors family. Biochemistry 28: 2392–2398.
15. Kratzner R, Debreczeni JE, Pape T, Schneider TR, Wentzel A, et al. (2005) Structure of Ecballium elaterium trypsin inhibitor II (EETI-II): a rigid molecular scaffold. Acta Crystallogr D Biol Crystallogr 61: 1255–1262.
16. Le-Nguyen D, Heitz A, Chiche L, el Hajji M, Castro B (1993) Characterization and 2D NMR study of the stable [9–21, 15–27] 2 disulfide intermediate in the folding of the 3 disulfide trypsin inhibitor EETI II. Protein Sci 2: 165–174.
17. Hernandez JF, Gagnon J, Chiche L, Nguyen TM, Andrieu JP, et al. (2000) Squash trypsin inhibitors from Momordica cochinchinensis exhibit an atypical macrocyclic structure. Biochemistry 39: 5722–5730.
18. Kimura RH, Cheng Z, Gambhir SS, Cochran JR (2009) Engineered knottin peptides: a new class of agents for imaging integrin expression in living subjects. Cancer Res 69: 2435–2442.
19. Werle M, Kafedjiiski K, Kolmar H, Bernkop-Schnürch A (2007) Evaluation and improvement of the properties of the novel cystine-knot microprotein McoEeTI for oral administration. Int J Pharm 332: 72–79.
20. Werle M, Schmitz T, Huang HL, Wentzel A, Kolmar H, et al. (2006) The potential of cystine-knot microproteins as novel pharmacophoric scaffolds in oral peptide drug delivery. J Drug Target 14: 137–146.
21. Jiang L, Kimura RH, Miao Z, Silverman AP, Ren G, et al. (2010) Evaluation of a ^{61}Cu-labeled cystine-knot peptide based on agouti-related protein for PET of tumors expressing $\alpha_v\beta_3$ integrin. J Nucl Med 51: 251–258.
22. Silverman AP, Levin AM, Lahti JL, Cochran JR (2009) Engineered cystine-knot peptides that bind $\alpha_v\beta_3$ integrin with antibody-like affinities. J Mol Biol 385: 1064–1075.
23. Reiss S, Sieber M, Oberle V, Wentzel A, Spangenberg P, et al. (2006) Inhibition of platelet aggregation by grafting RGD and KGD sequences on the structural scaffold of small disulfide-rich proteins. Platelets 17: 153–157.
24. Jiang L, Miao Z, Kimura RH, Silverman AP, Ren G, et al. (2012) ^{111}In-labeled cystine-knot peptides based on the Agouti-related protein for targeting tumor angiogenesis. J Biomed Biotechnol 2012: 368075.
25. Stricher F, Huang CC, Descours A, Duquesnoy S, Combes O, et al. (2008) Combinatorial optimization of a CD4-mimetic miniprotein and cocrystal structures with HIV-1 gp120 envelope glycoprotein. J Mol Biol 382: 510–524.
26. Vita C, Drakopoulou E, Vizzavona J, Rochette S, Martin L, et al. (1999) Rational engineering of a miniprotein that reproduces the core of the CD4 site interacting with HIV-1 envelope glycoprotein. Proc Natl Acad Sci U S A 96: 13091–13096.
27. Avrutina O, Schmoldt HU, Gabrijelcic-Geiger D, Wentzel A, Frauendorf H, et al. (2008) Head-to-tail cyclized cystine-knot peptides by a combined recombinant and chemical route of synthesis. Chembiochem 9: 33–37.
28. Sommerhoff CP, Avrutina O, Schmoldt HU, Gabrijelcic-Geiger D, Diederichsen U, et al. (2010) Engineered cystine knot miniproteins as potent inhibitors of human mast cell tryptase β. J Mol Biol 395: 167–175.
29. Christmann A, Walter K, Wentzel A, Kratzner R, Kolmar H (1999) The cystine knot of a squash-type protease inhibitor as a structural scaffold for Escherichia coli cell surface display of conformationally constrained peptides. Protein Eng 12: 797–806.
30. Wentzel A, Christmann A, Kratzner R, Kolmar H (1999) Sequence requirements of the GPNG β-turn of the Ecballium elaterium trypsin inhibitor II explored by combinatorial library screening. J Biol Chem 274: 21037–21043.
31. Getz JA, Rice JJ, Daugherty PS (2011) Protease-resistant peptide ligands from a knottin scaffold library. ACS Chem Biol 6: 837–844.
32. Thongyoo P, Jaulent AM, Tate EW, Leatherbarrow RJ (2007) Immobilized protease-assisted synthesis of engineered cysteine-knot microproteins. Chembiochem 8: 1107–1109.
33. Martin L, Stricher F, Misse D, Sironi F, Pugniere M, et al. (2003) Rational design of a CD4 mimic that inhibits HIV-1 entry and exposes cryptic neutralization epitopes. Nat Biotechnol 21: 71–76.
34. Reinwarth M, Glotzbach B, Tomaszowski M, Fabritz S, Avrutina O, et al. (2013) Oxidative folding of peptides with cystine-knot architectures: kinetic studies and optimization of folding conditions. Chembiochem 14: 137–146.
35. Takeuchi T, Shuman MA, Craik CS (1999) Reverse biochemistry: use of macromolecular protease inhibitors to dissect complex biological processes and identify a membrane-type serine protease in epithelial cancer and normal tissue. Proc Natl Acad Sci U S A 96: 11054–11061.
36. Yuan C, Chen L, Meehan EJ, Daly N, Craik DJ, et al. (2011) Structure of catalytic domain of Matriptase in complex with Sunflower trypsin inhibitor-1. BMC Struct Biol 11: 30.
37. Lin CY, Anders J, Johnson M, Sang QA, Dickson RB (1999) Molecular cloning of cDNA for matriptase, a matrix-degrading serine protease with trypsin-like activity. J Biol Chem 274: 18231–18236.
38. Oberst MD, Williams CA, Dickson RB, Johnson MD, Lin CY (2003) The activation of matriptase requires its noncatalytic domains, serine protease domain, and its cognate inhibitor. J Biol Chem 278: 26773–26779.
39. Lee MS, Tseng IC, Wang Y, Kiyomiya K, Johnson MD, et al. (2007) Autoactivation of matriptase in vitro: requirement for biomembrane and LDL receptor domain. Am J Physiol Cell Physiol 293: C95–105.
40. Tseng IC, Xu H, Chou FP, Li G, Vazzano AP, et al. (2010) Matriptase activation, an early cellular response to acidosis. J Biol Chem 285: 3261–3270.
41. List K, Haudenschild CC, Szabo R, Chen W, Wahl SM, et al. (2002) Matriptase/MT-SP1 is required for postnatal survival, epidermal barrier function, hair follicle development, and thymic homeostasis. Oncogene 21: 3765–3779.

Acknowledgments

We thank AB SCIEX Germany GmbH, Darmstadt for providing equipment and expertise in MS measurements of the cystine-knot peptides. We also thank Philipp Czechowski and Prof. Michael Reggelin for kindly providing equipment and measuring time for the CD spectra. This work was supported in part by Deutsche Forschungsgemeinschaft SPP1623 through grant KO1390/10-1.

Author Contributions

Conceived and designed the experiments: AC OA HK. Performed the experiments: BG MR NW SF MT HF. Analyzed the data: BG MR NW SF MT HF AC OA HK. Contributed reagents/materials/analysis tools: BG MR NW SF MT HF AC OA HK. Wrote the paper: BG MR AC OA HK.

42. Milner JM, Patel A, Davidson RK, Swingler TE, Desilets A, et al. (2010) Matriptase is a novel initiator of cartilage matrix degradation in osteoarthritis. Arthritis Rheum 62: 1955–1966.
43. Kilpatrick LM, Harris RL, Owen KA, Bass R, Ghorayeb C, et al. (2006) Initiation of plasminogen activation on the surface of monocytes expressing the type II transmembrane serine protease matriptase. Blood 108: 2616–2623.
44. Cheng MF, Jin JS, Wu HW, Chiang PC, Sheu LF, et al. (2007) Matriptase expression in the normal and neoplastic mast cells. Eur J Dermatol 17: 375–380.
45. Uhland K (2006) Matriptase and its putative role in cancer. Cell Mol Life Sci 63: 2968–2978.
46. List K (2009) Matriptase: a culprit in cancer? Future Oncol 5: 97–104.
47. Saleem M, Adhami VM, Zhong W, Longley BJ, Lin CY, et al. (2006) A novel biomarker for staging human prostate adenocarcinoma: overexpression of matriptase with concomitant loss of its inhibitor, hepatocyte growth factor activator inhibitor-1. Cancer Epidemiol Biomarkers Prev 15: 217–227.
48. Lee JW, Yong Song S, Choi JJ, Lee SJ, Kim BG, et al. (2005) Increased expression of matriptase is associated with histopathologic grades of cervical neoplasia. Hum Pathol 36: 626–633.
49. Avrutina O, Fittler H, Glotzbach B, Kolmar H, Empting M (2012) Between two worlds: a comparative study on in vitro and in silico inhibition of trypsin and matriptase by redox-stable SFTI-1 variants at near physiological pH. Org Biomol Chem.
50. Galkin AV, Mullen L, Fox WD, Brown J, Duncan D, et al. (2004) CVS-3983, a selective matriptase inhibitor, suppresses the growth of androgen independent prostate tumor xenografts. Prostate 61: 228–235.
51. Steinmetzer T, Schweinitz A, Sturzebecher A, Donnecke D, Uhland K, et al. (2006) Secondary amides of sulfonylated 3-amidinophenylalanine. New potent and selective inhibitors of matriptase. J Med Chem 49: 4116–4126.
52. Oberst MD, Chen LY, Kiyomiya K, Williams CA, Lee MS, et al. (2005) HAI-1 regulates activation and expression of matriptase, a membrane-bound serine protease. Am J Physiol Cell Physiol 289: C462–470.
53. Oberst MD, Johnson MD, Dickson RB, Lin CY, Singh B, et al. (2002) Expression of the serine protease matriptase and its inhibitor HAI-1 in epithelial ovarian cancer: correlation with clinical outcome and tumor clinicopathological parameters. Clin Cancer Res 8: 1101–1107.
54. List K, Szabo R, Molinolo A, Sriuranpong V, Redeye V, et al. (2005) Deregulated matriptase causes ras-independent multistage carcinogenesis and promotes ras-mediated malignant transformation. Genes Dev 19: 1934–1950.
55. Seitz I, Hess S, Schulz H, Eckl R, Busch G, et al. (2007) Membrane-type serine protease-1/matriptase induces interleukin-6 and -8 in endothelial cells by activation of protease-activated receptor-2: potential implications in atherosclerosis. Arterioscler Thromb Vasc Biol 27: 769–775.
56. Colombo É, Désilets A, Duchêne D, Chagnon F, Najmanovich R, et al. (2012) Design and synthesis of potent, selective inhibitors of matriptase. ACS Med Chem Let 3: 530–534.
57. Yamasaki Y, Satomi S, Murai N, Tsuzuki S, Fushiki T (2003) Inhibition of membrane-type serine protease 1/matriptase by natural and synthetic protease inhibitors. J Nutr Sci Vitaminol (Tokyo) 49: 27–32.
58. Boder ET, Wittrup KD (1997) Yeast surface display for screening combinatorial polypeptide libraries. Nat Biotechnol 15: 553–557.
59. Chao G, Lau WL, Hackel BJ, Sazinsky SL, Lippow SM, et al. (2006) Isolating and engineering human antibodies using yeast surface display. Nat Protoc 1: 755–768.
60. Desilets A, Longpre JM, Beaulieu ME, Leduc R (2006) Inhibition of human matriptase by eglin c variants. FEBS Lett 580: 2227–2232.
61. Van den Brulle J, Fischer M, Langmann T, Horn G, Waldmann T, et al. (2008) A novel solid phase technology for high-throughput gene synthesis. Biotechniques 45: 340–343.
62. Boy RG, Mier W, Nothelfer EM, Altmann A, Eisenhut M, et al. (2010) Sunflower trypsin inhibitor 1 derivatives as molecular scaffolds for the development of novel peptidic radiopharmaceuticals. Mol Imaging Biol 12: 377–385.
63. Morrison JF (1969) Kinetics of the reversible inhibition of enzyme-catalysed reactions by tight-binding inhibitors. Biochim Biophys Acta 185: 269–286.
64. Tischler M, Nasu D, Empting M, Schmelz S, Heinz DW, et al. (2012) Braces for the peptide backbone: insights into structure-activity relationships of protease inhibitor mimics with locked amide conformations. Angew Chem Int Ed Engl 51: 3708–3712.
65. Felizmenio-Quimio ME, Daly NL, Craik DJ (2001) Circular proteins in plants: solution structure of a novel macrocyclic trypsin inhibitor from Momordica cochinchinensis. J Biol Chem 276: 22875–22882.
66. Gera N, Hussain M, Rao BM (2012) Protein selection using yeast surface display. Methods.
67. Silverman AP, Levin AM, Lahti JL, Cochran JR (2009) Engineered cystine-knot peptides that bind $\alpha_v\beta_3$ integrin with antibody-like affinities. J Mol Biol 385: 1064–1075.
68. Silverman AP, Kariolis MS, Cochran JR (2011) Cystine-knot peptides engineered with specificities for $\alpha_{IIb}\beta_3$ or $\alpha_{IIb}\beta_3$ and $\alpha_v\beta_3$ integrins are potent inhibitors of platelet aggregation. J Mol Recognit 24: 127–135.
69. Chen M, Fu YY, Lin CY, Chen LM, Chai KX (2007) Prostasin induces protease-dependent and independent molecular changes in the human prostate carcinoma cell line PC-3. Biochim Biophys Acta 1773: 1133–1140.
70. Lahti JL, Silverman AP, Cochran JR (2009) Interrogating and predicting tolerated sequence diversity in protein folds: application to E. elaterium trypsin inhibitor-II cystine-knot miniprotein. PLoS Comput Biol 5: e1000499.
71. Chen J, Shiyanov P, Schlager JJ, Green KB (2012) A pseudo MS3 approach for identification of disulfide-bonded proteins: uncommon product ions and database search. J Am Soc Mass Spectrom 23: 225–243.
72. Suzuki M, Kobayashi H, Kanayama N, Saga Y, Lin CY, et al. (2004) Inhibition of tumor invasion by genomic down-regulation of matriptase through suppression of activation of receptor-bound pro-urokinase. J Biol Chem 279: 14899–14908.
73. Andreasen PA, Egelund R, Petersen HH (2000) The plasminogen activation system in tumor growth, invasion, and metastasis. Cell Mol Life Sci 57: 25–40.
74. Chen WH, Horoszewicz JS, Leong SS, Shimano T, Penetrante R, et al. (1982) Human pancreatic adenocarcinoma: in vitro and in vivo morphology of a new tumor line established from ascites. In Vitro 18: 24–34.
75. Tan MH, Shimano T, Chu TM (1981) Differential localization of human pancreas cancer-associated antigen and carcinoembryonic antigen in homologous pancreatic tumoral xenograft. J Natl Cancer Inst 67: 563–569.
76. Uhland K, Siphos B, Arkona C, Schuster M, Petri B, et al. (2009) Use of IHC and newly designed matriptase inhibitors to elucidate the role of matriptase in pancreatic ductal adenocarcinoma. Int J Oncol 35: 347–357.
77. Domoto T, Takino T, Guo L, Sato H (2012) Cleavage of hepatocyte growth factor activator inhibitor-1 by membrane-type MMP-1 activates matriptase. Cancer Sci 103: 448–454.
78. Quimbar P, Malik U, Sommerhoff CP, Kaas Q, Chan LY, et al. (2013) High-affinity cyclic peptide matriptase inhibitors. J. Biol. Chem. 288: 13885–13896.
79. Jiang L, Kimura RH, Miao Z, Silverman AP, Ren G, et al. (2010) Evaluation of a ^{64}Cu-labeled cystine-knot peptide based on agouti-related protein for PET of tumors expressing $\alpha_v\beta_3$ integrin. J Nucl Med 51: 251–258.
80. Ganesan R, Eigenbrot C, Kirchhofer D (2010) Structural and mechanistic insight into how antibodies inhibit serine proteases. Biochem J 430: 179–189.
81. LeBeau AM, Lee M, Murphy ST, Hann BC, Warren RS, et al. (2013) Imaging a functional tumorigenic biomarker in the transformed epithelium. Proc Natl Acad Sci U S A 110: 93–98.

3.3 Fragmentation follows Structure: Top-Down Mass Spectrometry Elucidates the Topology of Engineered Cystine-knot Miniproteins

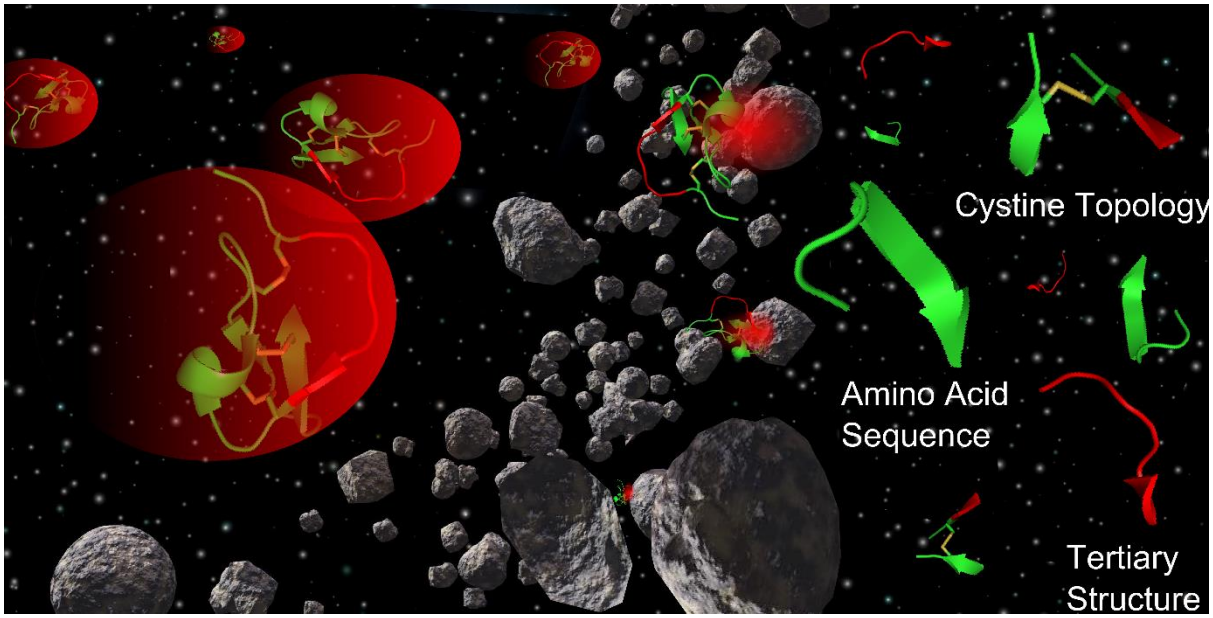
Title

Fragmentation Follows Structure: Top-Down Mass Spectrometry Elucidates the Topology of Engineered Cystine-Knot Miniproteins

Authors

Michael Reinwarth, Olga Avrutina, Sebastian Fabritz, Harald Kolmar

Graphical Abstract



Bibliographical Data

PLoS ONE 9(10): e108626.DOI: 10.1371/journal.pone.0108626

Received: 02 July 2014 / Accepted: 02 September 2014 / Published: 10 October 2014

Contributions by Michael Reinwarth

- Design of experiments
- Synthesis of all miniproteins
- Acquisition of mass-spectrometric data including fragmentation stages
- Interpretation of data
- Preparation of manuscript and figures

OPEN ACCESS Freely available online

PLOS ONE

Fragmentation Follows Structure: Top-Down Mass Spectrometry Elucidates the Topology of Engineered Cystine-Knot Miniproteins

Michael Reinwarth[1], Olga Avrutina[1], Sebastian Fabritz[2]*, Harald Kolmar[1]*

1 Institute of Organic Chemistry and Biochemistry, Technische Universität Darmstadt, Darmstadt, Germany, **2** AB Sciex Germany GmbH, Darmstadt, Germany

Abstract

Over the last decades the field of pharmaceutically relevant peptides has enormously expanded. Among them, several peptide families exist that contain three or more disulfide bonds. In this context, elucidation of the disulfide patterns is extremely important as these motifs are often prerequisites for folding, stability, and activity. An example of this structure-determining pattern is a cystine knot which comprises three constrained disulfide bonds and represents a core element in a vast number of mechanically interlocked peptidic structures possessing different biological activities. Herein, we present our studies on disulfide pattern determination and structure elucidation of cystine-knot miniproteins derived from *Momordica cochinchinensis* peptide MCoTI-II, which act as potent inhibitors of human matriptase-1. A top-down mass spectrometric analysis of the oxidised and bioactive peptides is described. Following the detailed sequencing of the peptide backbone, interpretation of the MS^3 spectra allowed for the verification of the knotted topology of the examined miniproteins. Moreover, we found that the fragmentation pattern depends on the knottin's folding state, hence, tertiary structure, which to our knowledge has not been described for a top-down MS approach before.

Citation: Reinwarth M, Avrutina O, Fabritz S, Kolmar H (2014) Fragmentation Follows Structure: Top-Down Mass Spectrometry Elucidates the Topology of Engineered Cystine-Knot Miniproteins. PLoS ONE 9(10): e108626. doi:10.1371/journal.pone.0108626

Editor: Scheherazade Sadegh-Nasseri, Johns Hopkins University, United States of America

Received July 2, 2014; **Accepted** September 2, 2014; **Published** October 10, 2014

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files.

Funding: The authors have no support or funding to report.

Competing Interests: S. Fabritz is associated with a commercial company, AB Sciex Germany GmbH. There are no patents, products in development or marketed products to declare. All other authors declare no competing interests. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* Email: Sebastian.Fabritz@absciex.com (SF); Kolmar@Biochemie-TUD.de (HK)

Introduction

Precise information concerning identity, structure, and pharmacokinetics of drug candidates is an important issue in the development of biopharmaceuticals, among them a vast number of bioactive peptides [1]. Therefore, reliable and reproducible analysis of their structure and topology needs well-elaborated high-throughput methods. Tandem mass spectrometry (MS/MS or MS^2) has become a valuable tool for the identification and quantification of peptides and proteins. Recent experiments indicate that MS analysis can also be applied to their structural characterization since folded and unfolded molecules may give rise to different fragmentation patterns upon ionization [1–3]. Moreover, MS^2-based methods can be applied to identify and characterize inter- and intramolecular disulfide bonds along with NMR and X-ray analysis [2]. Following the bottom-up methodology for the determination of disulfide-bond topologies in multi-disulfide proteins, the analytes are partially reduced and subjected to enzymatic digestion prior to the analysis *via* MS or MS/MS [4–6]. Although this approach provides decisive information on the primary structure, determination of disulfide connectivities *via* MS analysis of proteolytic fragments of non-reduced or partially reduced peptides with high disulfide content remains a complicated task.

In some instances, the available specific enzymatic cleavage sites within the peptide of interest do not necessarily lead to one-cystine-one-peptide fragment distributions after digestion of folded species, or no appropriate cleavage sites are present at all. Moreover, partial reduction and S-alkylation prior to proteolysis often results in a complex mixture of variants [4,7,8]. Furthermore, lack of information originating from incomplete sequence coverage and the fact that not every posttranslational modification (PTM) could be detected, are additional disadvantages [4,5,9–15].

In recent years, efforts have been made to apply a top-down methodology for the characterization of full-length native proteins and peptides *via* concurrent cleavage of disulfide bonds using MS/MS methods minimizing loss of information [8,11–14,16–18]. To this end, different MS fragmentation methods were applied, among them the widely employed collision-induced dissociation (CID) featuring low collision energies. Electron-transfer or capture dissociation (ETD or ECD, respectively) methods have also been used for the generation of MS/MS spectra [8,17–21]. For CID the mobile proton theory, which states that positive charges are randomly distributed among all amino acids of an analyzed peptide, thus facilitating amide N-protonation and cleavage, is an important tool to interprete MS spectra [16,22–27]. Regarding arginine-rich peptides this is not entirely true since the basic side chains are known to sequester charges [16,22–24]. Thus, the ratio

of charge to the number of arginines within the sequence must be over one to make the mobile proton theory applicable [16,22–24]. Hence, the overall charge of the peptide plays an important role since protonated arginine side chains are known to promote disulfide cleavage [15,19–24]. As a consequence, the collision energy required for the cleavage of disulfides depends on the ratio of cystine to arginine units within the sequence [16,22–24]. Additional difficulties upon assigning peptidic fragments in the complex spectra progressively increase with the size of the peptide and the number of S-S bonds [8,17,18]. The side chains of several amino acids tend to neutral losses and the formation of uncommon product ions under ionization conditions due to specific proximity effects [2,8,17,18,28]. In particular, asymmetric cleavage of disulfide bonds results in the formation of both perthiocysteine (Ptc) and dehydroalanine (Dha), of which the latter is known to induce cleavage of the peptide backbone not at the amide bond, but promotes fragmentation between the amide nitrogen and the α-carbon, hence forming c-ions (Figure 1C) [2,8,16,28,29]. These effects lead to an increased number of signals in the spectrum resulting in a statistically high chance of false-positive assignments. Thus, regarding large disulfide-rich peptides and proteins, the gained spectra require further in-depth analysis due to their complexity [30].

Following previous studies on the analysis of cysteine-rich proteins, Green and co-workers elaborated the idea of an additional fragmentation stage (MS^3) and used it for a top-down approach to analyze chicken lysozyme [2,22,31,32]. Therein, nozzle-skimmer ionization was used as the first dissociation source, while regular CID served as a second fragmentation stage [2]. Although this pseudo MS^3 approach has delivered promising results, the MS^3 utilizing two dissociation sources could facilitate the detailed evaluation of the disulfide pattern topology.

As model disulfide-rich macromolecules for investigation of structural features and disulfide bond connectivities by application of MS^3 methodologies, engineered cystine-knot protease inhibitors were used in the present study. These bioactive peptides, also known as knottins, consist of about 30 amino acid residues and are characterized by a unique, 'pseudo-knotted' architecture [33–35].

Knottin's structural core is defined by three β-strands which are interconnected by three disulfide bonds. The bonds between CysII and CysV as well as CysI and CysIV form a ring which is penetrated by a third cystine, connecting CysIII and CysVI (Figure 1B) [34–36]. Additional thermodynamic stability is provided *via* an extensive network of hydrogen bonds between the peptide backbone and the side chains of the amino acids located within the β–sheets [37,38]. Therefore peptides containing this structural motif display an exceptional structural, thermal, and biological robustness [35,39,40].

Knottins are considered excellent scaffolds for the generation of tailor-made compounds for diagnostic and therapeutic applications since the surface-exposed loops consist of segments with high structural and numeral flexibility, while the conserved core only tolerates minor amino acid exchanges [32,35,41–43]. Engineered miniproteins derived from trypsin inhibitors of the bitter gourd *Momordica cochinchinensis* (MCoTI) have been successfully applied for the inhibition of proteinases of clinical relevance, among them human mast cell tryptase β, foot-and-mouth-disease virus 3C protease, and, most recently, cancer-related protease matriptase-1 (Figure 1D) [44–46]. Moreover, a cystine-knot peptide from cone snails, Ziconotide, has been approved for the treatment of severe and chronic pain [47]. Although optimized conditions for both chemical and recombinant synthesis of these miniproteins have been recently reported, oxidative folding towards the respective bioactive isomer still reveals no information on the disulfide topology [7,48,49]. Biological activity is an important indication for a correct fold, but for engineered variants with multiple amino acid exchanges this information may not necessarily prove the knotted cystine pattern [45].

To date, determination of disulfide bond connectivities for this class of peptides relies on NMR and X-Ray of native peptides as well as on MS analysis of chemically or enzymatically modified ones [17,18]. Compared to the NMR and X-Ray technology, a major advantage of an analytical technique such as a mass spectrometry is the requirement of less analyte. Despite the fact that method development may be time consuming, the broad availability of MS instruments enable more laboratories to make structural determinations and, when using well-established methods, analysis can be performed in shorter time. Thus, Tam and coworkers recently described disulfide mapping of the cyclotide hedyotide B2 by MS^2 analysis of the partially reduced and alkylated peptide [50]. Furthermore, Balaram et al. elucidated disulfide topology of four-cysteine native peptides as well as enzymatically treated conotoxines [17]. Camadro *et al*. described structural analysis of knotted Psalmopeotoxin I through combined tandem MS and ^{15}N–NMR [18]. In the present study we applied MS^3 technology for the characterization of folded and oxidised open-chain MCoTI variants. For the confirmation of structural and topological characteristics we employed a top-down approach to deliver information on elements regarding the primary, secondary, and tertiary structure of the miniprotein.

Materials and Methods

Peptides **1**, **2**, and **3** have been synthesized as previously reported [44,49].

Detailed information on mass spectrometric measurements is provided in the supporting information (Method S1). Briefly, the 4000 QTRAP LC/MS/MS and the 6500 QTRAP LC/MS/MS systems (AB Sciex Germany GmbH, Table S1) were used. For MS full scans experiments signal intensity was adjusted to achieve 1e6 counts per second (cps) maximal signal intensity at 10 µL per minute infusion from a syringe pump. Peptide stock solutions have been prepared by dissolving 0.5 mg of the respective peptide in 1 mL of a mixture of 20% methanol (Fluka Analytical, LC-MS CHROMASOLV) and 80% water (Merck KGaA, LC-MS LiChrosolv) with 0.2% formic acid (Fluka Analytical, analytic additive).

For MS^2 experiments both declustering potential (DP) and collision energy (CE) have been optimized *via* ramping. Regarding MS^3 measurements, resonance excitation energy (AF2) was additionally optimized (using 'ramping' functionality of Analyst software). All spectra were accumulated to ensure distinct differentiation between signals and noise.

Results

Experimental design and general results

As model peptides we chose variants derived from the cystine-knot peptide MCoTI (Figure 1), a synthetic protease inhibitor combining the three-disulfide pattern with the open-chain amide backbone [44]. Since peptides of this family are known for their structural stability regarding biological and chemical decomposition, denaturation or disulfide scrambling should not take place under acidic conditions which were used for mass-spectrometric analysis [7]. The examined miniprotein variants comprised selected exchanges within the flexible loops [44,49].

To investigate the influence of structural characteristics on the fragmentation pattern, tandem mass spectrometric analysis triple

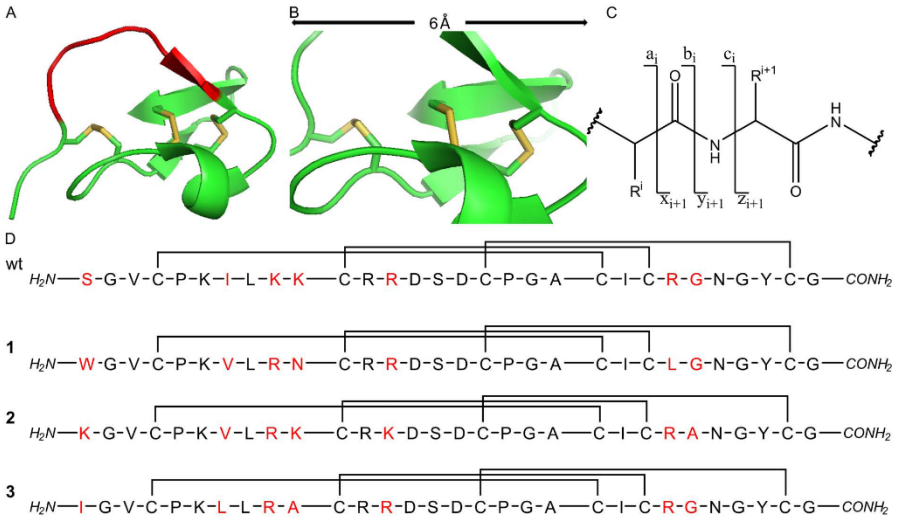


Figure 1. Structure of synthetic open-chain MCoTI, structural overview and fragment ion formation. (**A**) 3D structure of synthetic open-chain cystine knot oMCoTI (pdb: 2IT8).15 Active loop is shown in red. (**B**) A 6 Å close-up on the disulfide-tightened core of MCoTI. For both (**A**) and (**B**) disulfides are shown in yellow. (**C**) Overview of the generation of peptidic fragment ions upon CID. (**D**) Sequences of the parent MCoTI wild type (wt) and the miniproteins (**1–3**) used in this study. Positions with altered amino acid residues (regarding the wild type sequence) are marked red.
doi:10.1371/journal.pone.0108626.g001

quadrupole systems with a linear ion trap were used (ESI, Table S1). The design of these systems allows for two dissociation steps upon selection of the respective precursor ions. Thus, we applied CID to both the oxidised three-disulfide peptide 1 (Figure 1) and its reduced precursor. In the resulting spectra, the unique peaks and those showing the typical shift of hydrogens or sulfhydryl groups (Figure 2 and Figures S1–S8) were compared.

Identification of the respective fragment was achieved *via* selection of the corresponding precursor for MS^3 (Figure 2 and Figures S9–S13). Ion path parameters and ionization energies, namely the declustering potential, the collision energy and the excitation energy for the MS^3 were semi-automatically optimized (ramped) towards maximum intensity of the respective product ions. Similar fragmentation patterns were found for peptides 1, 2, and 3 that deviated from each other in the amino acid sequences of loop 1, 2, and 5 (Figure 1) [44].

While amino acid exchanges outside the inhibitory loop 1 that protrudes into the active site of the respective protease induced no altered fragmentation of the native peptide, as expected, relative intensities of the respective ions dropped upon arginine exchanges since protonated arginine side chains are known to promote disulfide cleavage [15,19–21]. Indeed, the relative peak intensity of the C-terminal fragments (y_{14}, Figure 2 and Figures S7 and S8) is higher for the arginine-containing variants (**2** and **3**) compared to **1** that lacks this moiety. In contrast, for the fragment $c_{16}y_{19}$ or $c_{16}z_{20}$, respectively, only a minor decrease of signal intensity was observed for peptide **3**, since it still contained an additional arginine which was kept unchanged in each variant. Moreover, signal intensity of product ion series immediately drops upon cleavage of arginine, which is caused by the sequestered charge at the arginine side chains, thus decreasing the intensity of the following b-ions by leaving the majority of them neutral (Figures 3B and Figure S9).

The MS^2 spectra only showed minor differences between the unfolded and folded variants (Figure 2). Indeed, the fragmentation patterns were similar, apart from the expected unsymmetrical cystine dissociations resulting in Ptc or Dha moieties and different intensities for the fragments $c_{16}y_{19}$ ($c_{16}z_{20}$) and, particularly, $c_{10}y_{26}$.

CID spectra obtained from the lower charged precursors $[M+2H]^{2+}$ or $[M+3H]^{3+}$, respectively, were significantly enriched with three fragment species (c_{10}, $c_{16}y_{19}$ and y_{14}) covering the whole sequence (Figure 2). These major fragments were identified and sequenced using MS^3 methodology. This pattern lacked Cys11, which was caused by the formation of the major c-ion fragment c_{10} (Figure 2). C-ion formation has been reported to occur upon cysteine fragmentation towards Dha [2].

For high-charged states, e.g. $[M+5H]^{5+}$, MS^2 spectra of the folded miniprotein displayed a different fragmentation pattern. Interestingly, we observed formation of several b-ions which could not be detected upon low-charge CID (Figure 2). While all b-ions comprising large linear fragments still possessed at least one disulfide, one fragment was identified lacking the central SD motif, thus resulting in the generation of a 'bridged' fragment (Figure 3D). However, assignment of the fragments was complicated by the multiple charge states within the spectra.

Elucidation of primary structure

For the sequential analysis, the major fragment ions c_{10}, $c_{16}y_{19}$ and y_{14} were chosen for MS^3 analysis. Assignment of b- and y-ions was complicated since several obviously predetermined breaking points dictated the shape of the spectra, while other fragments were hard to distinguish from noise (Figure 3). Additionally, 'satellite' peaks were observed for each sub-fragment ion,

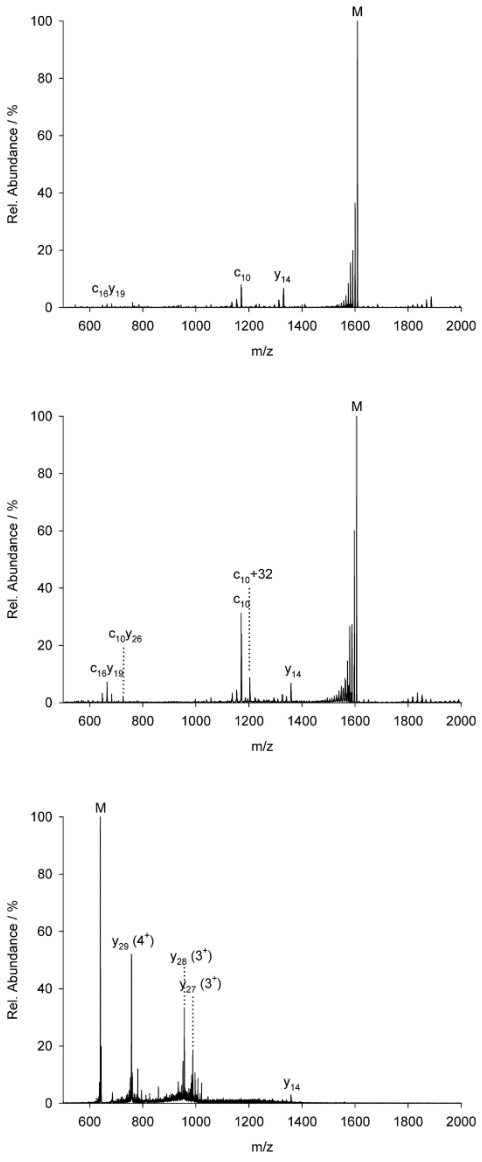


Figure 2. MS^2 of MCoTI peptide 1. (**A**) CID of the doubly charged ion of reduced peptide 1. (**B**) CID of the doubly charged ion of folded miniprotein. (**C**) CID of the five-fold charged ion of folded miniprotein.
doi:10.1371/journal.pone.0108626.g002

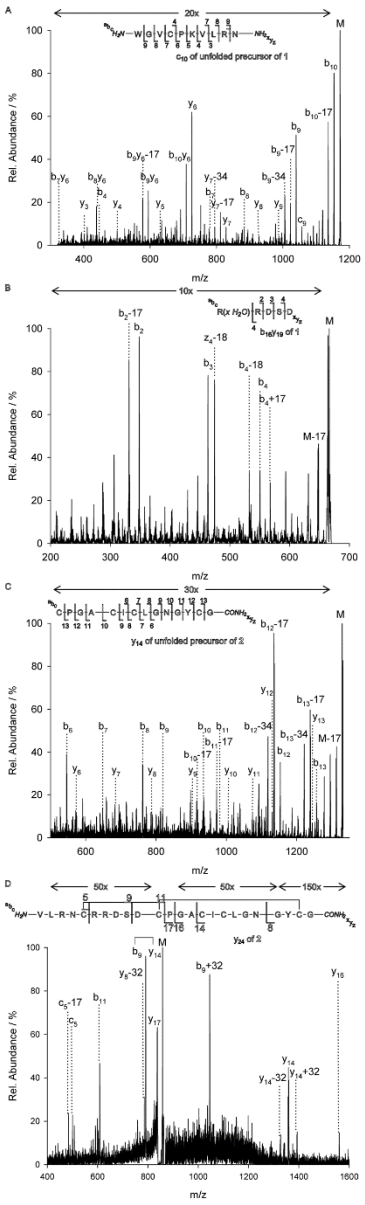


Figure 3. MS^3 of the CID-obtained major fragments of MCoTI 1. (**A**) MS^3 of c_{10} of the doubly charged ions of unfolded peptide. (**B**) MS^3 of $c_{16}y_{19}$ of the doubly-charged ions of folded miniprotein. (**C**) MS^3 of y_{14} of the doubly charged ions of unfolded peptide. (**D**) MS^3 of y_{24} of the five-fold charged ions of folded miniprotein. Arrows above the spectra indicate intensity amplifications.
doi:10.1371/journal.pone.0108626.g003

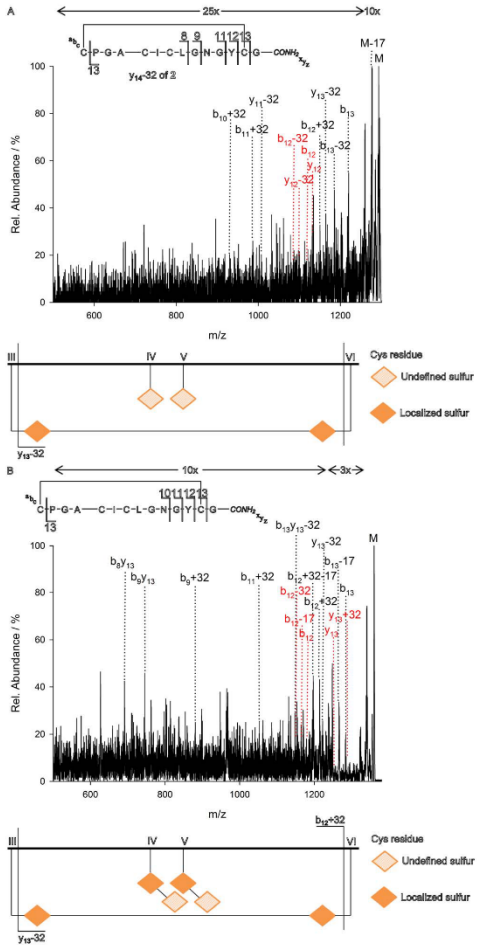


Figure 4. MS^3 of y_{14}-32 (A) and y_{14}+32 (B) and the resulting combinatorial interpretation. In red are inexistent peaks to provide evidence on the respective Ptc or Dha cleavage. Arrows above the spectra indicate intensity amplifications.
doi:10.1371/journal.pone.0108626.g004

originated from neutral losses of amino acid side chains, e. g. lysine (−17 Da) [51].

In the case of the aminoterminal fragment c_{10}, intensity of b- and c-ions decreased upon cleavage of Arg9 (which is equivalent to the formation of the b_9/c_9 ion), due to the sequestered charge at the guanidyl moiety (Figure 3A). Regarding the y-ions, the dominating fragment in the MS^3 spectrum was the y_6-ion which became the starting point for further b- ion fragmentation leading to the generation of $b_{10}y_6$, b_9y_6, b_8y_6, and b_7y_6 ions (Figure 3A). Moreover, the y-ion series was detected until formation of y_3.

The $c_{16}y_{19}$, respectively, $c_{16}z_{20}$, fragment was analyzed by MS^3 and sequencing revealed its water adduct. For this fragment we

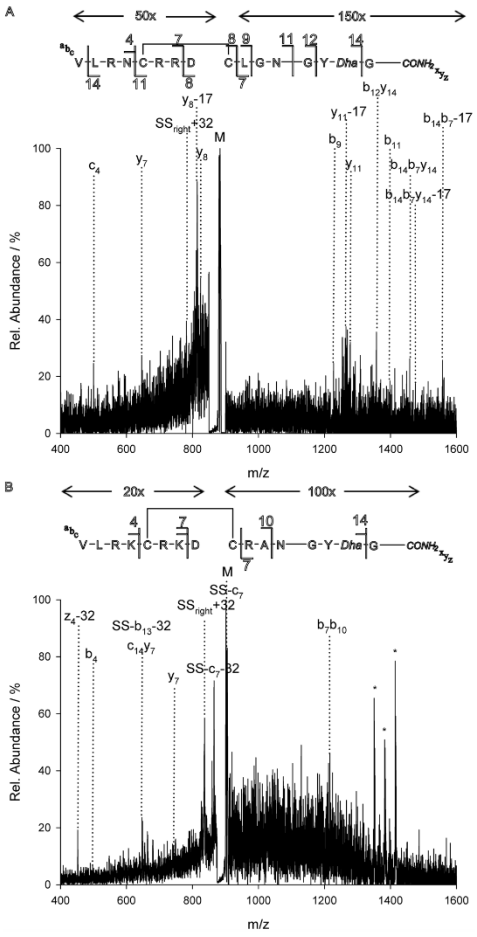


Figure 5. Spectra and assigned structures of the bridged fragments. (**A**) MS^3 of the bridged fragment of **1**. (**B**) MS^3 of the bridged fragment of **2**. Arrows above the spectra indicate intensity amplifications.
doi:10.1371/journal.pone.0108626.g005

observed b_4 as well as y_4, y_3, and y_2-subfragments which provided detailed sequence information (Figure 3B). As expected, the spectrum was enriched by y-ions formed upon aspartate cleavages. Interestingly, the b_4 ion fragment was not observed as water adduct anymore, thus indicating Arg12 being primarily responsible for the binding of water (Figure 3B).

The carboxyterminal fragment y_{14} delivered complete information regarding the primary structure. B-ion formation from b_{13} to b_6, as well as y-ion formation of y_{13} to y_6 (Figure 3C) was observed. However, this spectrum was also complicated by predetermined breaking points as well as neutral losses, due to several proximity effects or directed non-standard fragmentation like Dha-promoted c-ion formation.

Table 1. Amount of loop ratios in CID of the respective peptides.

	Intensity of $c_{10}y_{26}$ (% of TIC)		Ratio	Intensity of $c_{16}y_{19}$ (% of TIC)		Ratio
	Reduced	**Folded**		**Reduced**	**Folded**	
1 (CE: 70 V)	0.25	2.2	1:9	1.1	7.2	1:7
2 (CE: 50 V)	0.17	3.2	1:19	<0.1	1.8	1:18
3 (CE: 85 V)	3.5	21.5	1:6	1.8	9.8	1:5

doi:10.1371/journal.pone.0108626.t001

Secondary structure

Determination of the cystine connections could not be achieved directly through detection of 'bridged' fragments for lower-charge states. This is due to the fact that the cysteine residues are close to each other within the sequence, e. g. having only one isoleucine in between CysIV and CysV. Moreover, analysis is complicated by the intrinsic knotted structure and the fact that disulfide dissociation occurs upon charge localization on basic side chains [8,17,18].

For this reason the four-cysteine fragment y_{14} was of particular interest as it contained an intact disulfide bridge. For the determination of the expected III–VI connection, we analyzed three related m/z fragments, namely, 1294, 1326 and 1358, *via* MS^3. These ions correspond to a four-cysteine-three-sulfur (y_{14}-32, Figure 4A), a four-cysteine-four-sulfur ($y_{14,}$ Figure 3C) and a four-cysteine-five-sulfur (y_{14}+32, Figure 4B) fragment, respectively. MS^3 analysis of those fragments of **1** suffered from the absence of an arginine compared to **2**, thus these fragment series were displayed a higher intensity in the arginine-containing variant (Figures S11 and S12). Nevertheless, combinatorial interpretation of the resulting spectra clearly revealed the predicted connection of CysIII to CysVI (Figure 4). Involvement of CysIII was verified since it was found to be cut off exclusively as a Ptc moiety for all three parent fragments (Figure 4A), hence the remaining y ions were exclusively found as y-32 ions. After cleavage of the Ptc from the four-cysteine-three-sulfur fragment, a three-cysteine-one-sulfur one (y_{13}-32, Figure 4A) was received, while the cleavage of Cys (y_{13}) or Dha (y_{13}+32) moieties did not occur (Figure 4A and Figures S11 and S12, red 'peaks'). Indeed, the distribution of the sulfur atoms was exclusively found to be on CysIV and CysV. Hence, connections of IV–V, IV–VI, and V–VI could be excluded. As a result, the only possible intramolecular connection was the expected III–VI one. To provide further evidence of the connection CysIII to CysVI, analysis of the four-cysteine-five-sulfur fragment was performed. After generation of the b_{12} ion, a three-cysteine-three-sulfur fragment (b_{12}+32, Figure 4B) was obtained, which is underpinned through the absence of the ions b_{12}-32 and b_{12} (Figure 4B and Figures S11 and S12, red 'peaks'). Thus, connections of III–IV and III–V could be excluded solely leaving III–VI one possible. Nonetheless, regarding the lower charged precursors, no exact conclusion could be made on the two remaining cystines, which still could be the native, knotted I–IV/II–V or the intuitively implausible, unknotted I–V/II–IV connection.

For verification of the knottin fold, the second disulfide connection was evaluated. However, the probability that a bridged fragment is formed upon CID is very low since at least three peptide bonds need to be cleaved while the respective cystine connection has to remain intact. For this purpose we focused on the higher charged states (5+) for the relationship between the protonated arginine residues and the fragmentation of disulfide bonds (see introduction section) [16,23,24].

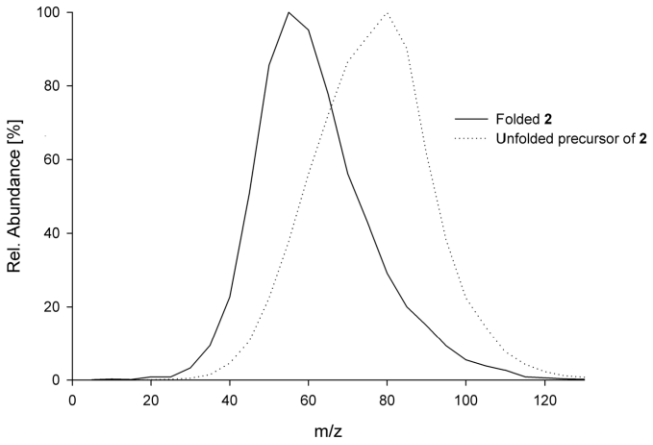


Figure 6. Optimization of the collision energy towards a maximum formation of fragment $c_{10}y_{26}$ of 2.
doi:10.1371/journal.pone.0108626.g006

However, we found the linear y_{24} fragment containing the connections CysIII–CysVI as well as CysII–CysIV or CysII–CysV, respectively (Figure 3D). Upon cleavage of the labile SD motif, this fragment comprised two linear fragments connected through one intermolecular cystine. For the proof of the II–V connection, this fragment was subjected to MS^3 analysis. This procedure revealed an ion consisting of two linear peptide chains connected through the disulfide linkage between CysII and CysV. Direct MS^3 analysis of this bridged fragment produced several ions belonging to the expected product connected *via* the linkage of CysII and CysV (Figure 5). Since peak intensities were relatively low due to the unlikeliness of the multiple, specific breakages necessary for the generation of this fragment, the complete fragment series could not be detected although further optimizations of the detector parameters could increase the signal to noise ratio. Nevertheless, the major product ions were identified.

For peptide **2** this procedure of fragment analysis was interfered by another fragment ion with nearly identical mass (Figure 5), which might be avoided by tightening the isolation window of the precursor ions. Nevertheless, the expected product ion was verified, too. Additionally, fragmentation revealed a small fragment ion which was identified to originate from the II–V-connected peptide ion (Figure S13). This in-depth analysis was not performed for verification of peptide **3**.

Tertiary structure

Comparison of the CID spectra of the positively charged folded and unfolded peptides revealed a distinct intensity increase of the peak corresponding to both the $c_{16}y_{19}$ as well as the $c_{10}y_{26}$ fragment for the folded miniprotein (Figure 2), which was confirmed by MS^3 analysis. Generation of both fragments was compared for each miniprotein variant at different collision energies. Ratios depicted in Table 1 result from collision energies optimized towards a maximum TIC.

For miniproteins **1** and **3** ratios of over 6 regarding the formation of both fragments for the reduced peptide and the folded knottin were observed. For miniprotein **2** the observed ratio was approximately 20. Additionally, the dependency of these ratios as well as the maximum fragment formation on the applied collision energy in obedience to the folding state was elucidated. Interestingly, for **1** and **3** no influence of the folding state on the maximum fragment formation was observed. For miniprotein **2**, instead, increasing collision energies lead to an adjustment in fragment formation. Thus, a semi-automatic optimization of the collision energy to determine the energy with a maximum formation of fragment $c_{10}y_{26}$ (Figure 6) was performed. Maxima at different collision energies were observed for the reduced peptide and the folded knottin. This accounted for high ratios at CE 50 as well as decreasing ratios for increasing collision energies.

Discussion

The analyzed miniprotein variants displayed very similar fragmentation patterns, notwithstanding the fact that sequence alterations have been applied for all sequential and structural regions (Figure 1C). Interestingly, the resulting fragmentation pattern was not influenced by sequential changes. Therefore, the reproducibility of the results in this experimental setting should allow for the prediction of CID spectra for novel representatives of the MCoTI family.

Fragmentation was obviously guided by cysteines as well as aspartates. In non-mobile proton systems aspartate side chains are known to induce carboxyterminal cleavage [2,16]. Hence, predetermined cleavage sites within the "DSD" motif of the respective peptide variants were verified. Regarding the cysteine residues, the fragmentation pattern was predetermined by CysII and CysIII of the sequence, although the folding state, hence secondary and tertiary structural elements, only had influence on signal intensity rather than fragment formation itself. This indicates a cysteine-induced cleavage of the peptide backbone resulting in the formation of c-ions for both reduced and oxidised cysteines [2,16]. In contrast, the charge state played a key role in fragmentation. Higher charged precursors led to the formation of large b ion fragments. As a consequence, only a decreased amount of cysteine-guided fragmentation could be observed. Both findings strongly correlate with the fragmentation patterns reported by Green and co-workers [2,16].

We achieved full sequence coverage for the reduced and oxidised peptides in this top-down approach. For all three peptides similar fragments were detectable and could be identified. Hence, for evidence of known sequences or those with minor exchanges, sequence analysis can be performed easily without sophisticated methods of sample preparation or spectra analysis.

Disulfide topology is the major determinant of secondary structure for cystine-knot miniproteins. Since no predetermined breaking points resulting in 'bridged' fragments were found in CID-based spectra, in-depth analysis towards determination of the cystine pattern was performed for relatively large, linear fragments containing intramolecular disulfides. MS^3-based sequencing was performed and combinatorial interpretation of the resulting spectra revealed the expected disulfide connectivity (Figure 4). This analysis was based on the generation of Dha or Ptc moieties upon CID resulting from unsymmetrical breakage of cystines. This untypical manner of analysis has rarely been reported for the topology determination in a top-down approach since it is hardly applicable for large proteins [17,18,52]. The specific restrictions present for cystine-knot miniproteins, in particular, the short sequential distances between the respective cysteine residues and the overlap of cystine connections, almost exclude the generation of 'bridged' fragments which are not connected *via* the amide backbone and thus the exact determination of disulfide patterns [17,18].

Interpretation of the spectra of the bridged fragments for the determination of a second cystine connection was complicated through their complex nature and low intensities, thus intensive manual analysis was required. Since multiple collisions are possible within the collision cell and the linear ion trap, the analyzed fragments indeed most likely were *tertiary* or *quaternary* fragment ions. Moreover, probabilities for the generation of those multiple fragmentation events are relatively low, peak intensities of the resulting 'pseudo-MS^n' spectra are very low as well. This only allowed for detection of sequentially predetermined breaking points, thus no formation of fragment ion series was observed. Instead, several individual species with increased intensities were determined and the bridged fragments could be clearly identified for both **1** and **2**, although the spectra of **2** were interfered by another fragment ion with a similar mass (Figure 5). Despite this complicated methodology for the *ab initio* determination of disulfide patterns, verification of knotted disulfide connectivity of oxidised peptides of the MCoTI family now can be routinely applied without time-consuming methods by fragment analysis of common CID spectra.

We observed increased intensities of the fragments corresponding to the surface-exposed loops in the CID spectra of the oxidatively folded miniproteins. Since some loops of knottins of the MCoTI family are subjected to tensed conformations upon oxidative folding, their increased energetic content may account for the increased peak intensities (Table 1). This is particularly

reasonable for loop 1 (flanked by CysI and CysII) as its biological function, i.e. binding to the respective protease, strongly depends on the folding state of the peptide, which requires the correct disulfide pattern along with the surface-exposed and tensed inhibitory loop region [53]. Accordingly, these elements of the tertiary structure most likely account for the rise in peak intensity.

Further interest was spent on the investigation of the shifted maxima in collision energy regarding the formation of fragment $c_{10}y_{26}$ of reduced and folded **2**. Interestingly, peptides **1** and **3** did not show this behavior. The additional basic amino acid Lys10 in the inhibitory loop 1 as well as a slightly different 3D structure may account for this interesting characteristic of **2**. Applying this type of analysis, determination of the correct cystine-knot fold of miniproteins, at least with sequences similar to **1**, can now be achieved through comparison of fragment formation in the folded and unfolded state. MS-based detection of the influence of chemically modified amino acids on the tertiary structure of proteins has often been reported, but to the best of our knowledge, an effect of the tertiary structure on the fragmentation pattern has not been described for a top-down approach before [1,8,54–56].

Conclusion

In this study a set of synthetic matriptase-1 inhibitors were analyzed through MS^3 methodology. Full sequence coverage and determination of disulfide topology were achieved. Determination of primary structure is of particular interest for 'one-bead-one-compound' libraries which have already been applied for linear peptides and for medium throughput screening of bioactive molecules [57,58]. It would be interesting to see whether this methodology can be applied to the on-bead sequence analysis of cystine-knot peptides although oxidative folding of the resin-bound peptides is still critical [59]. Additionally, an influence of secondary and tertiary structure on the fragmentation pattern was observed. As a consequence, the folding state and thus the knotted nature of the disulfides can easily be verified by detecting either formation of the respective fragments upon CID or *via* comparison of the ratios of the formation of loop 1 in the folded and unfolded state, which should be typically in the range of 10. Taking into consideration the reproducibility of this analysis, this methodology can be applied to the analysis of structure and disulfide bond connectivities of other miniproteins of the MCoTI family and may be useful for other peptides with related multidisulfide pattern and interesting pharmacological properties as e.g. cyclotides or conotoxines [17,18]. The influence of tertiary structure on fragmentation patterns of MS/MS spectra may lead to information on the region of protein-protein interaction with ramifications for analysis of various biomolecular interactions as e.g. antibody-antigen binding or conformational changes of antigens upon antibody binding.

Supporting Information

Figure S1 Initial spectra of 1. (**A**) MS^1 of **1.** (**B**) Zoom-In on $M+2H]^{2+}$. (**C**) Zoom-In on $M+5H]^{5+}$.
(PNG)

Figure S2 Initial spectra of unfolded precursor of 1. (**A**) MS^1 of unfolded precursor of **1.** (**B**) Zoom-In on $M+2H]^{2+}$.
(PNG)

Figure S3 Initial spectra of 2. (**A**) MS^1 of **2.** (**B**) Zoom-In on $M+2H]^{2+}$. (**C**) Zoom-In on $M+5H]^{5+}$.
(PNG)

Figure S4 Initial spectra of unfolded precursor of 2. (**A**) MS^1 of unfolded precursor of **2.** (**B**) Zoom-In on $M+2H]^{2+}$.
(PNG)

Figure S5 Initial spectra of 3. (**A**) MS^1 of **3.** (**B**) Zoom-In on $M+2H]^{2+}$.
(PNG)

Figure S6 Initial spectra of unfolded precursor of 3. (**A**) MS^1 of unfolded precursor of **3.** (**B**) Zoom-In on $M+2H]^{2+}$.
(PNG)

Figure S7 MS^2 of MCoTI peptide 2. (**A**) CID of the triply charged ion of reduced precursor of peptide **2.** (**B**) CID of the triply charged ion of folded miniprotein **2.** (**C**) CID of the fivefold charged ion of folded miniprotein **2.**
(PNG)

Figure S8 MS^2 of MCoTI peptide 3. (**A**) CID of the triply charged ion of reduced precursor of peptide **3.** (**B**) CID of the triply charged ion of folded miniprotein **3.**
(PNG)

Figure S9 MS^3 of major fragments of MCoTI peptide 2. (**A**) MS^3 of c_{10} ion. (**B**) MS^3 of $b_{16}z_{20}$ ion. (**C**) MS^3 of y_{14} ion. Arrows above the spectra indicate intensity amplifications.
(PNG)

Figure S10 MS^3 of major fragments of MCoTI peptide 3. (**A**) MS^3 of c_{10} ion. (**B**) MS^3 of y_{14} ion. Arrows above the spectra indicate intensity amplifications.
(PNG)

Figure S11 MS^3 of y_{14}-32 (A) and y_{14}+32 (B) for the combinatorial interpretation of 2. In red are inexistent peaks to provide evidence on the respective Ptc or Dha cleavage. Arrows above the spectra indicate intensity amplifications.
(PNG)

Figure S12 MS^3 of y_{14}-32 (A) and y_{14}+32 (B) for the combinatorial interpretation of 3. In red are inexistent peaks to provide evidence on the respective Ptc or Dha cleavage. Asterisk indicates fragment from different parent ion with identical mass. Arrows above the spectra indicate intensity amplifications.
(PNG)

Figure S13 MS^3 of $b_{14}z_{20}$-32 fragment of MCoTI peptide 2.
(PNG)

Table S1 Applied source and gas parameters.
(DOCX)

Method S1 Experimental procedure for mass spectrometric measurements.
(DOC)

Acknowledgments

We thank Iain Gibb and Dietrich Merkel (AB SCIEX) for their steady support of this project. AB SCIEX Germany GmbH, Darmstadt, generally provided equipment and expertise in MS measurements of the cystine-knot peptides.

Author Contributions

Conceived and designed the experiments: SF MR OA HK. Performed the experiments: MR SF. Analyzed the data: MR SF OA HK. Contributed reagents/materials/analysis tools: MR SF OA HK. Contributed to the writing of the manuscript: MR SF OA HK.

References

1. Berkowitz SA, Engen JR, Mazzeo JR, Jones GB (2012) Analytical tools for characterizing biopharmaceuticals and the implications for biosimilars. Nat Rev Drug Discov 11: 527–540.
2. Chen J, Shiyanov P, Schlager JJ, Green KB (2012) A pseudo MS3 approach for identification of disulfide-bonded proteins: uncommon product ions and database search. J Am Soc Mass Spectrom 23: 225–243.
3. Seidler J, Zinn N, Boehm ME, Lehmann WD (2010) De novo sequencing of peptides by MS/MS. Proteomics 10: 634–649.
4. Goransson U, Craik DJ (2003) Disulfide mapping of the cyclotide kalata B1. Chemical proof of the cystic cystine knot motif. J Biol Chem 278: 48188–48196.
5. Gundry RL, White MY, Murray CI, Kane LA, Fu Q, et al. (2009) Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow. Curr Protoc Mol Biol Chapter 10: Unit10 25.
6. Switzar L, Giera M, Niessen WM (2013) Protein digestion: an overview of the available techniques and recent developments. J Proteome Res 12: 1067–1077.
7. Reinwarth M, Nasu D, Kolmar H, Avrutina O (2012) Chemical synthesis, backbone cyclization and oxidative folding of cystine-knot peptides: promising scaffolds for applications in drug design. Molecules 17: 12533–12552.
8. Goyder MS, Rebeaud F, Pfeifer ME, Kalman F (2013) Strategies in mass spectrometry for the assignment of Cys-Cys disulfide connectivities in proteins. Expert Rev Proteomics 10: 489–501.
9. Castagnola M, Cabras T, Iavarone F, Vincenzoni F, Vitali A, et al. (2012) Top-down platform for deciphering the human salivary proteome. J Matern Fetal Neonatal Med 25: 27–43.
10. Kellie JF, Tran JC, Lee JE, Ahlf DR, Thomas HM, et al. (2010) The emerging process of Top Down mass spectrometry for protein analysis: biomarkers, protein-therapeutics, and achieving high throughput. Mol Biosyst 6: 1532–1539.
11. Lanucara F, Eyers CE (2013) Top-down mass spectrometry for the analysis of combinatorial post-translational modifications. Mass Spectrom Rev 32: 27–42.
12. Macht M (2009) Mass spectrometric top-down analysis of proteins. Bioanalysis 1: 1131–1148.
13. Peng Y, Yu D, Gregorich Z, Chen X, Beyer AM, et al. (2013) In-depth proteomic analysis of human tropomyosin by top-down mass spectrometry. J Muscle Res Cell Motil.
14. Chait BT (2006) Chemistry. Mass spectrometry: bottom-up or top-down? Science 314: 65–66.
15. Zhang H, Ge Y (2011) Comprehensive analysis of protein modifications by top-down mass spectrometry. Circ Cardiovasc Genet 4: 711.
16. Chen J, Shiyanov P, Zhang L, Schlager JJ, Green-Church KB (2010) Top-down characterization of a native highly intralinked protein: concurrent cleavages of disulfide and protein backbone bonds. Anal Chem 82: 6079–6089.
17. Gupta K, Kumar M, Balaram P (2010) Disulfide bond assignments by mass spectrometry of native natural peptides: cysteine pairing in disulfide bonded conotoxins. Anal Chem 82: 8313–8319.
18. Combes A, Choi SJ, Pimentel C, Darbon H, Waidelich D, et al. (2009) Determination with matrix-assisted laser desorption/ionization tandem time-of-flight mass spectrometry of the extensive disulfide bonding in tarantula venom peptide Psalmopeotoxin I. Eur J Mass Spectrom (Chichester, Eng) 15: 517–529.
19. Chrisman PA, Pitteri SJ, Hogan JM, McLuckey SA (2005) SO2-* electron transfer ion/ion reactions with disulfide linked polypeptide ions. J Am Soc Mass Spectrom 16: 1020–1030.
20. Guthals A, Bandeira N (2012) Peptide identification by tandem mass spectrometry with alternate fragmentation modes. Mol Cell Proteomics 11: 550–557.
21. Zubarev RA, Kruger NA, Fridriksson EK, Lewis MA, Horn DM, et al. (1999) Electron Capture Dissociation of Gaseous Multiply-Charged Proteins Is Favored at Disulfide Bonds and Other Sites of High Hydrogen Atom Affinity. J Am Chem Soc 121: 2857–2862.
22. Ginter JM, Zhou F, Johnston MV (2004) Generating protein sequence tags by combining cone and conventional collision induced dissociation in a quadrupole time-of-flight mass spectrometer. J Am Soc Mass Spectrom 15: 1478–1486.
23. Lioe H, O'Hair RA (2007) A novel salt bridge mechanism highlights the need for nonmobile proton conditions to promote disulfide bond cleavage in protonated peptides under low-energy collisional activation. J Am Soc Mass Spectrom 18: 1109–1123.
24. Wells JM, Stephenson Jr JL, McLuckey SA (2000) Charge dependence of protonated insulin decompositions. Int J Mass Spec 203: A1–A9.
25. Wysocki VH, Tsaprailis G, Smith LL, Breci LA (2000) Mobile and localized protons: a framework for understanding peptide dissociation. J Mass Spectrom 35: 1399–1406.
26. Boyd R, Somogyi A (2010) The mobile proton hypothesis in fragmentation of protonated peptides: a perspective. J Am Soc Mass Spectrom 21: 1275–1278.
27. Cheng C, Gross ML (2000) Applications and mechanisms of charge-remote fragmentation. Mass Spectrom Rev 19: 398–420.
28. Paizs B, Suhai S (2005) Fragmentation pathways of protonated peptides. Mass Spectrom Rev 24: 508–548.
29. Chrisman PA, McLuckey SA (2002) Dissociations of disulfide-linked gaseous polypeptide/protein anions: ion chemistry with implications for protein identification and characterization. J Proteome Res 1: 549–557.
30. Wu SL, Jiang H, Lu Q, Dai S, Hancock WS, et al. (2009) Mass spectrometric determination of disulfide linkages in recombinant therapeutic proteins using online LC-MS with electron-transfer dissociation. Anal Chem 81: 112–122.
31. Körner R, Wilm M, Morand K, Schubert M, Mann M (1996) Nano electrospray combined with a quadrupole ion trap for the analysis of peptides and protein digests. J Am Soc Mass Spectrom 7: 150–156.
32. Raska CS, Parker CE, Huang C, Han J, Glish GL, et al. (2002) Pseudo-MS3 in a MALDI orthogonal quadrupole-time of flight mass spectrometer. J Am Soc Mass Spectrom 13: 1034–1041.
33. Craik DJ, Daly NL, Waine C (2001) The cystine knot motif in toxins and implications for drug design. Toxicon 39: 43–60.
34. Gelly JC, Gracy J, Kaas Q, Le-Nguyen D, Heitz A, et al. (2004) The KNOTTIN website and database: a new information system dedicated to the knottin scaffold. Nucleic Acids Res 32: D156–159.
35. Gracy J, Le-Nguyen D, Gelly JC, Kaas Q, Heitz A, et al. (2008) KNOTTIN: the knottin or inhibitor cystine knot scaffold in 2007. Nucleic Acids Res 36: D314–319.
36. Chiche L, Heitz A, Gelly JC, Gracy J, Chau PT, et al. (2004) Squash inhibitors: from structural motifs to macrocyclic knottins. Curr Protein Pept Sci 5: 341–349.
37. Felizmenio-Quimio ME, Daly NL, Craik DJ (2001) Circular proteins in plants: solution structure of a novel macrocyclic trypsin inhibitor from Momordica cochinchinensis. J Biol Chem 276: 22875–22882.
38. Heitz A, Hernandez JF, Gagnon J, Hong TT, Pham TT, et al. (2001) Solution structure of the squash trypsin inhibitor MCoTI-II. A new family for cyclic knottins. Biochemistry 40: 7973–7983.
39. Craik DJ, Cemazar M, Wang CK, Daly NL (2006) The cyclotide family of circular miniproteins: nature's combinatorial peptide template. Biopolymers 84: 250–266.
40. Wang CK, Hu SH, Martin JL, Sjogren T, Hajdu J, et al. (2009) Combined X-ray and NMR analysis of the stability of the cyclotide cystine knot fold that underpins its insecticidal activity and potential use as a drug scaffold. J Biol Chem 284: 10672–10683.
41. Kimura RH, Cheng Z, Gambhir SS, Cochran JR (2009) Engineered knottin peptides: a new class of agents for imaging integrin expression in living subjects. Cancer Res 69: 2435–2442.
42. Werle M, Kafedjiiski K, Kolmar H, Bernkop-Schnurch A (2007) Evaluation and improvement of the properties of the novel cystine-knot microprotein McoEeTI for oral administration. Int J Pharm 332: 72–79.
43. Werle M, Schmitz T, Huang HL, Wentzel A, Kolmar H, et al. (2006) The potential of cystine-knot microproteins as novel pharmacophoric scaffolds in oral peptide drug delivery. J Drug Target 14: 137–146.
44. Glotzbach B, Reinwarth M, Weber N, Fabritz S, Tomaszowski M, et al. (2013) Combinatorial Optimization of Cystine-Knot Peptides towards High-Affinity Inhibitors of Human Matriptase-1. PLoS One 8: e76956.
45. Sommerhoff CP, Avrutina O, Schmoldt HU, Gabrijelcic-Geiger D, Diederichsen U, et al. (2010) Engineered cystine knot miniproteins as potent inhibitors of human mast cell tryptase beta. J Mol Biol 395: 167–175.
46. Thongyoo P, Roque-Rosell N, Leatherbarrow RJ, Tate EW (2008) Chemical and biomimetic total syntheses of natural and engineered MCoTI cyclotides. Org Biomol Chem 6: 1462–1470.
47. Klotz U (2006) Ziconotide–a novel neuron-specific calcium channel blocker for the intrathecal treatment of severe chronic pain–a short review. Int J Clin Pharmacol Ther 44: 478–483.
48. Hofmeyer T, Bulani SI, Grzeschik J, Krah S, Glotzbach B, et al. (2013) Protein Production in Yarrowia lipolytica Via Fusion to the Secreted Lipase Lip2p. Mol Biotechnol.
49. Reinwarth M, Glotzbach B, Tomaszowski M, Fabritz S, Avrutina O, et al. (2013) Oxidative folding of peptides with cystine-knot architectures: kinetic studies and optimization of folding conditions. Chembiochem 14: 137–146.
50. Nguyen GK, Zhang S, Wang W, Wong CT, Nguyen NT, et al. (2011) Discovery of a linear cyclotide from the bracelet subfamily and its disulfide mapping by top-down mass spectrometry. J Biol Chem 286: 44833–44844.
51. Papayannopoulos IA (1995) The interpretation of collision-induced dissociation tandem mass spectra of peptides. Mass Spectrometry Reviews 14: 49–73.
52. Choi S, Jeong J, Na S, Lee HS, Kim HY, et al. (2010) New algorithm for the identification of intact disulfide linkages based on fragmentation characteristics in tandem mass spectra. J Proteome Res 9: 626–635.
53. Heitz A, Avrutina O, Le-Nguyen D, Diederichsen U, Hernandez JF, et al. (2008) Knottin cyclization: impact on structure and dynamics. BMC Struct Biol 8: 54.
54. Pan J, Borchers CH (2013) Top-down structural analysis of posttranslationally modified proteins by Fourier transform ion cyclotron resonance-MS with hydrogen/deuterium exchange and electron capture dissociation. Proteomics 13: 974–981.
55. Engen JR (2009) Analysis of protein conformation and dynamics by hydrogen/deuterium exchange MS. Anal Chem 81: 7870–7875.
56. Heck AJR (2008) Native mass spectrometry: a bridge between interactomics and structural biology. Nat Meth 5: 927–933.
57. Aina OH, Liu R, Sutcliffe JL, Marik J, Pan CX, et al. (2007) From combinatorial chemistry to cancer-targeting peptides. Mol Pharm 4: 631–651.

58. Lam KS, Liu R, Miyamoto S, Lehman AL, Tuscano JM (2003) Applications of one-bead one-compound combinatorial libraries and chemical microarrays in signal transduction research. Acc Chem Res 36: 370–377.

59. Atherton E, Clive DL, Sheppard RC (1975) Letter: Polyamide supports for polypeptide synthesis. J Am Chem Soc 97: 6584–6585.

4 Ergänzende Informationen zu den vorgelegten Publikation aus 3

4.1 Oxidative Folding of Peptides with Cystine-knot Architechture: Kinetic studies and optimization of folding conditions.

CHEMBIOCHEM

Supporting Information

Oxidative Folding of Peptides with Cystine-Knot Architectures: Kinetic Studies and Optimization of Folding Conditions

Michael Reinwarth, Bernhard Glotzbach, Michael Tomaszowski, Sebastian Fabritz, Olga Avrutina,* and Harald Kolmar*[a]

cbic_201200604_sm_miscellaneous_information.pdf

Analytical Data

Table S1. Overview of linear precursors

Precursor for miniprotein	Retention time [min]	Calculated masses	Measured masses	Yield (crude, according to resin loading)
1	13.12[a]	$M+2H]^{2+}$: 1624.41	$M+2H]^{2+}$: 1624.16	248.7 mg
		$M+3H]^{3+}$: 1083.27	$M+3H]^{3+}$: 1083.03	76.6 %
		$M+4H]^{4+}$: 812.70	$M+4H]^{4+}$: 812.55	
		$M+5H]^{5+}$: 650.36	$M+5H]^{5+}$: 650.24	
2	12.56[a]	$M+3H]^{3+}$: 1091.61	$M+3H]^{3+}$: 1091.78	226.4 mg
		$M+4H]^{4+}$: 818.96	$M+4H]^{4+}$: 818.69	69.2 %
		$M+5H]^{5+}$: 655.37	$M+5H]^{5+}$: 655.61	
3	14.19[a]	$M+3H]^{3+}$: 1098.98	$M+3H]^{3+}$: 1099.19	253.4
		$M+4H]^{4+}$: 824.49	$M+4H]^{4+}$: 824.71	76.9 %
		$M+5H]^{5+}$: 659.79	$M+5H]^{5+}$: 660.02	
4	12.77[a]	$M+3H]^{3+}$: 1067.94	$M+3H]^{3+}$: 1067.95	186.1
		$M+4H]^{4+}$: 801.21	$M+4H]^{4+}$: 801.40	58.1 %
		$M+5H]^{5+}$: 641.16	$M+5H]^{5+}$: 641.35	
5	16.00[a]	$M+3H]^{3+}$: 1072.92	$M+3H]^{3+}$: 1072.49	216.4
		$M+4H]^{4+}$: 804.94	$M+4H]^{4+}$: 804.70	67.3 %
		$M+5H]^{5+}$: 644.15	$M+5H]^{5+}$: 643.95	
6	13.63[a]	$M+3H]^{3+}$: 1101.94	$M+3H]^{3+}$: 1102.13	234.9
		$M+4H]^{4+}$: 826.71	$M+4H]^{4+}$: 826.90	71.1 %
		$M+5H]^{5+}$: 661.57	$M+5H]^{5+}$: 661.83	
7	18.63[b]	$M+3H]^{3+}$: 1144.36	$M+3H]^{3+}$: 1144.67	128.3
		$M+4H]^{4+}$: 858.52	$M+4H]^{4+}$: 858.76	37.4 %
8	14.49[b]	$M+3H]^{3+}$: 1177.70	$M+3H]^{3+}$: 1177.88	157.0
		$M+4H]^{4+}$: 883.52	$M+4H]^{4+}$: 883.66	44.5 %
		$M+5H]^{5+}$: 707.02	$M+5H]^{5+}$: 707.28	
9	14.87[b]	$M+3H]^{3+}$: 1280.14	$M+3H]^{3+}$: 1279.83	316.2
		$M+4H]^{4+}$: 960.36	$M+4H]^{4+}$: 960.07	82.4 %
		$M+5H]^{5+}$: 768.49	$M+5H]^{5+}$: 769.06	
10	14.61[b]	$M+3H]^{3+}$: 1254.78	$M+3H]^{3+}$: 1255.04	178.1
		$M+4H]^{4+}$: 941.33	$M+4H]^{4+}$: 941.51	47.3 %
		$M+5H]^{5+}$: 753.27	$M+5H]^{5+}$: 753.49	
		$M+6H]^{6+}$: 627.89	$M+6H]^{6+}$: 628.17	
11	17.71[b]	$M+2H]^{2+}$: 1565.91[c]	$M+2H]^{2+}$: 1566.48	124.4
		$M+3H]^{3+}$: 1044.27[c]	$M+3H]^{3+}$: 1044.47	39.7 %
		$M+4H]^{4+}$: 783.45[c]	$M+4H]^{4+}$: 783.74	
12	14.59[b]	$M+4H]^{4+}$: 1630.37	$M+4H]^{4+}$: 1630.29	20.9 mg[d]
		$M+5H]^{5+}$: 1304.50	$M+5H]^{5+}$: 1304.48	74.9 %
		$M+6H]^{6+}$: 1087.25	$M+6H]^{6+}$: 1087.18	
		$M+7H]^{7+}$: 932.07	$M+7H]^{7+}$: 932.00	
		$M+8H]^{8+}$: 815.69	$M+8H]^{8+}$: 815.62	
		$M+9H]^{9+}$: 725.17	$M+9H]^{9+}$: 725.09	
		$M+10H]^{10+}$: 652.75	$M+10H]^{10+}$: 652.70	
		$M+11H]^{11+}$: 593.50	$M+11H]^{11+}$: 593.44	

[a] Linear gradient of eluent B from 10 to 60 % over 20 minutes. [b] Linear gradient of eluent B from 10 to 80 % over 20 minutes. [c] Dithiopropiolic acid is coupled aminoterminal, which is partly cleaved during folding process. [d] Aprotinin **12** was reduced from its pure and native form.

Table S2. Overview of folded and eventually functionalized miniproteins.

Miniprotein	Retention time [min]	Calculated masses	Measured masses	Yield (pure, according to crude product)
1	9.68 [a]	$M+3H]^{3+}$: 1081.26	$M+3H]^{3+}$: 1081.40	8.5 mg
		$M+4H]^{4+}$: 811.19	$M+4H]^{4+}$: 811.45	22.0 %
		$M+5H]^{5+}$: 649.15	$M+5H]^{5+}$: 649.30	
1 ox	11.23 [a]	$M+3H]^{3+}$: 1071.23	$M+3H]^{3+}$: 1070.80	16.0 mg
		$M+H_2O+3H]^{3+}$: 1077.24	$M+H_2O+3H]^{3+}$: 1077.24	10.4 %
		$M+4H]^{4+}$: 803.67	$M+4H]^{4+}$: 803.53	
		$M+H_2O+4H]^{4+}$: 808.18	$M+H_2O+4H]^{4+}$: 808.18	
		$M+5H]^{5+}$: 643.14	$M+5H]^{5+}$: 642.91	
		$M+H_2O+5H]^{5+}$: 646.75	$M+H_2O+5H]^{5+}$: 646.75	
2	9.49 [a]	$M+3H]^{3+}$: 1089.59*	$M+3H]^{3+}$: 1089.86	6.4 mg
		$M+4H]^{4+}$: 817.45	$M+4H]^{4+}$: 817.74	19.1 %
		$M+5H]^{5+}$: 654.16	$M+5H]^{5+}$: 654.45	
3	11.29 [a]	$M+3H]^{3+}$: 1096.97	$M+3H]^{3+}$: 1097.22	11.7 mg
		$M+4H]^{4+}$: 822.98	$M+4H]^{4+}$: 823.26	27.8 %
		$M+5H]^{5+}$: 658.58	$M+5H]^{5+}$: 658.86	
4	9.39 [a]	$M+3H]^{3+}$: 1065.93	$M+3H]^{3+}$: 1066.14	7.2 mg
		$M+4H]^{4+}$: 799.70	$M+4H]^{4+}$: 799.93	18.2 %
		$M+5H]^{5+}$: 639.96	$M+5H]^{5+}$: 640.23	
5	13.61 [a]	$M+3H]^{3+}$: 1070.90	$M+3H]^{3+}$: 1071.00	5.8 mg
		$M+4H]^{4+}$: 803.43	$M+4H]^{4+}$: 803.56	14.1 %
		$M+5H]^{5+}$: 642.94	$M+5H]^{5+}$: 643.09	
6	11.98 [a]	$M+3H]^{3+}$: 1099.93	$M+3H]^{3+}$: 1099.03	4.2 mg
		$M+4H]^{4+}$: 825.20	$M+4H]^{4+}$: 824.59	11.5 %
		$M+5H]^{5+}$: 660.36	$M+5H]^{5+}$: 659.98	
6 ox	12.28[a]	$M+2H]^{2+}$: 1642.88	$M+2H]^{2+}$: 1642.93	7.4 mg
		$M+3H]^{3+}$: 1095.58	$M+3H]^{3+}$: 1095.59	6.3 %
		$M+4H]^{4+}$: 821.94	$M+4H]^{4+}$: 821.97	
7	16.78 [b]	$M+3H]^{3+}$: 1142.34	$M+3H]^{3+}$: 1142.38	3.9 mg
		$M+4H]^{4+}$: 857.01	$M+4H]^{4+}$: 857.07	9.6 %
8	14.04 [b]	$M+3H]^{3+}$: 1175.68	$M+3H]^{3+}$: 1175.73	3.6 mg
		$M+4H]^{4+}$: 882.01	$M+4H]^{4+}$: 882.08	9.5 %
		$M+5H]^{5+}$: 705.81	$M+5H]^{5+}$: 705.94	
9	13.92 [b]	$M+3H]^{3+}$: 1278.13	$M+3H]^{3+}$: 1278.11	10.8 mg
		$M+4H]^{4+}$: 958.85	$M+4H]^{4+}$: 958.77	23.8 %
10	13.14 [b]	$M+3H]^{3+}$: 1252.76	$M+3H]^{3+}$: 1252.71	6.8 mg
		$M+4H]^{4+}$: 939.82	$M+4H]^{4+}$: 939.84	16.2 %
11	14.78 [b]	$M+2H]^{3+}$: 1510.82	$M+2H]^{3+}$: 1510.94	1.5 mg
		$M+3H]^{4+}$: 1007.55	$M+3H]^{4+}$: 1007.59	7.0 %
12	10.17 [b]	$M+4H]^{4+}$: 1628.86	$M+4H]^{4+}$: 1628.78	1.2 mg
		$M+5H]^{5+}$: 1303.29	$M+5H]^{5+}$: 1303.26	13.3 %
		$M+6H]^{6+}$: 1086.24	$M+6H]^{6+}$: 1086.14	
		$M+7H]^{7+}$: 931.24	$M+7H]^{7+}$: 931.11	

[a] Linear gradient of eluent B from 10 to 60 % over 20 minutes. [b] Linear gradient of eluent B from 10 to 80 % over 20 minutes.

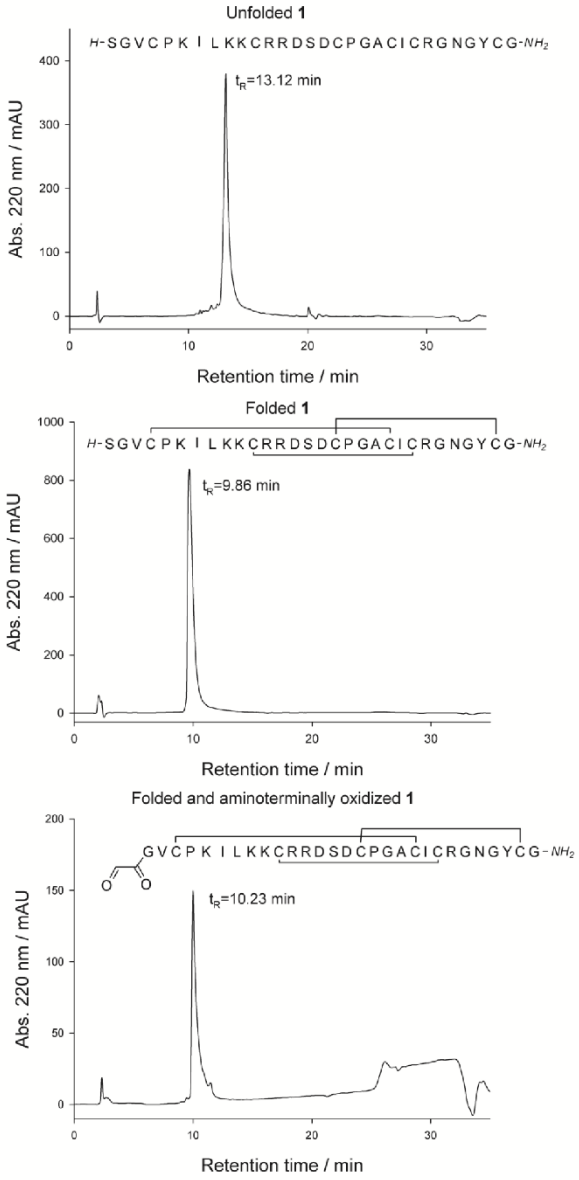


Figure S1. HPLC trace at 220 nm (10→60) of **1** in unfolded, folded and folded and aldehydic state.

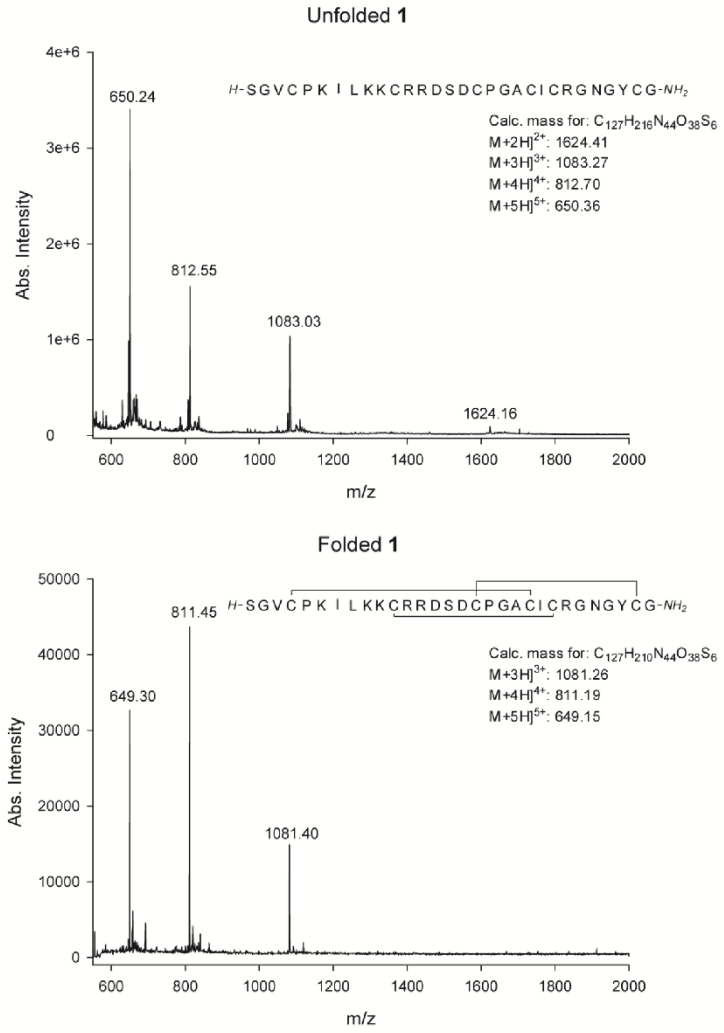


Figure S2. ESI-MS spectra of **1** in unfolded and folded state.

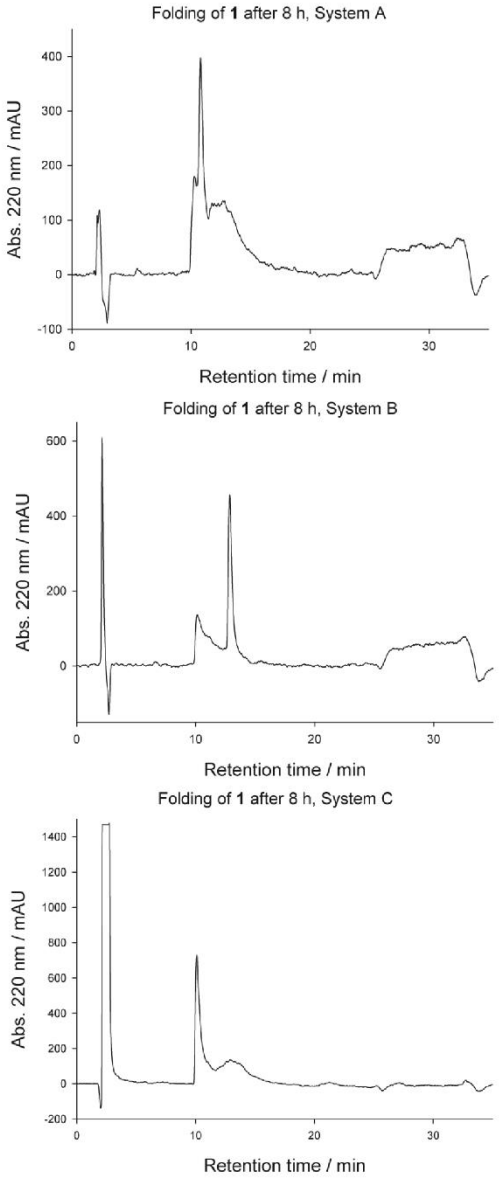


Figure S3. HPLC trace at 220 nm (10→60) of **1** in the corresponding folding mixture after 8 h.

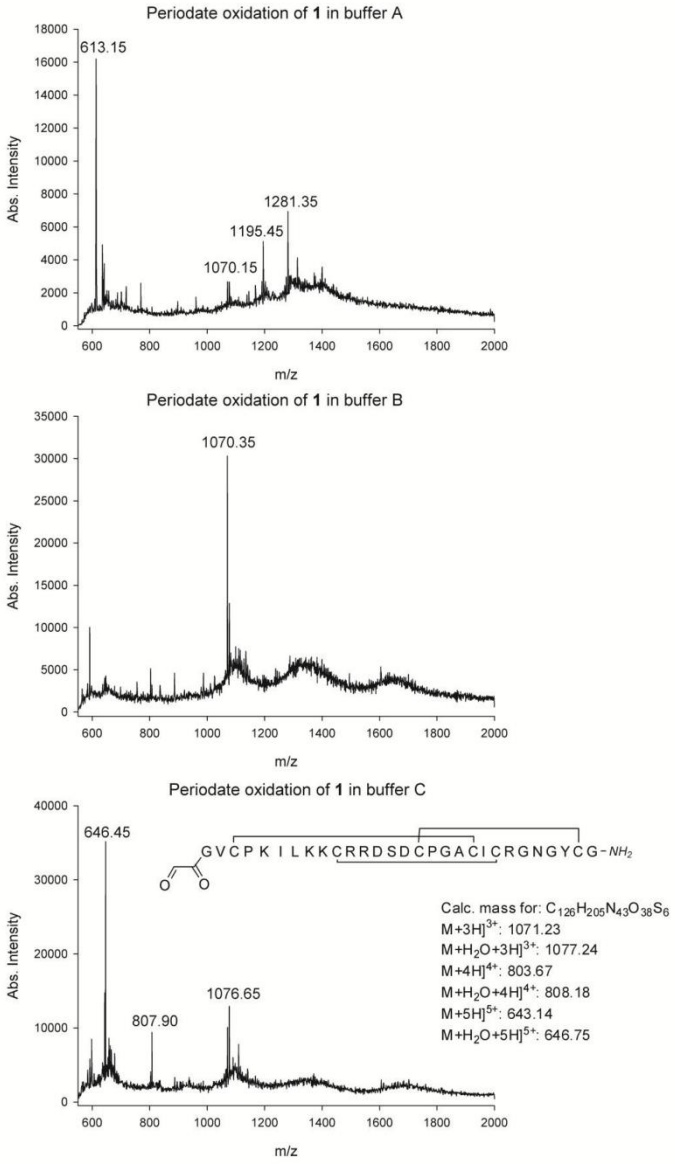


Figure S4. ESI-MS spectra of periodate oxidation of **1** in the folding mixtures after 10 min.

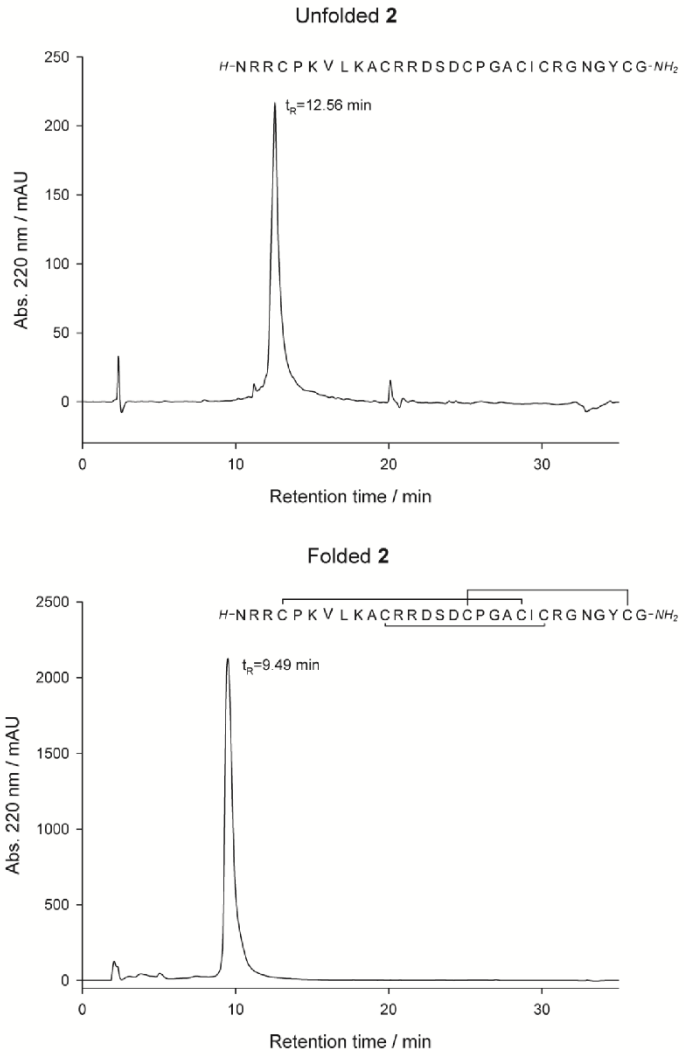


Figure S SEQ Fig._ * ARABIC 5. HPLC trace at 220 nm (10→60) of **2** in unfolded and folded state.

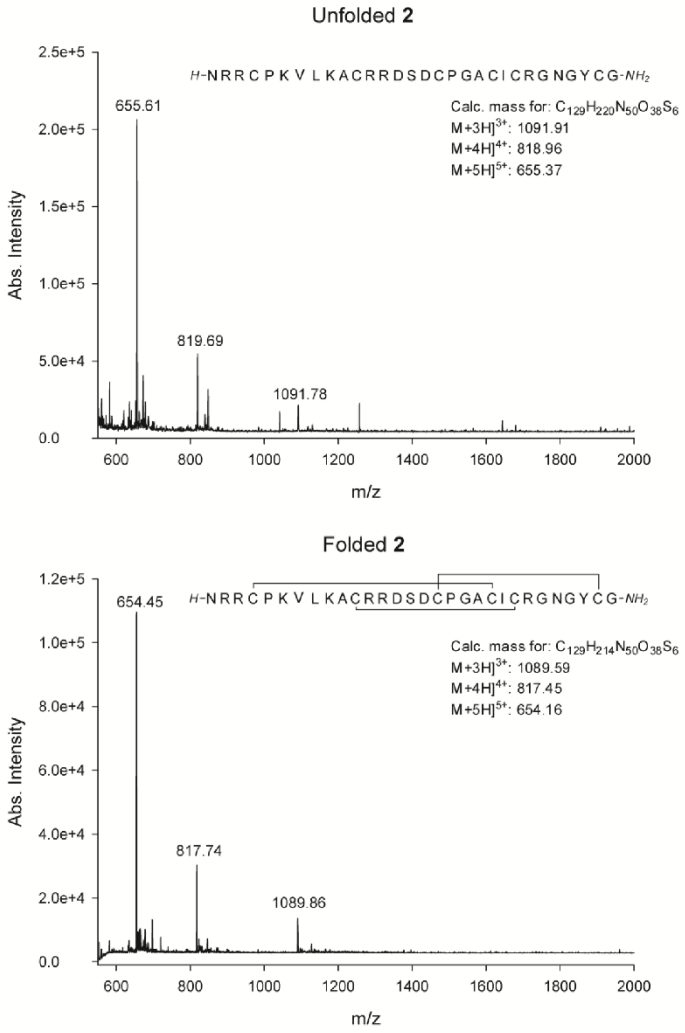


Figure S5. ESI-MS spectra of **2** in unfolded and folded state.

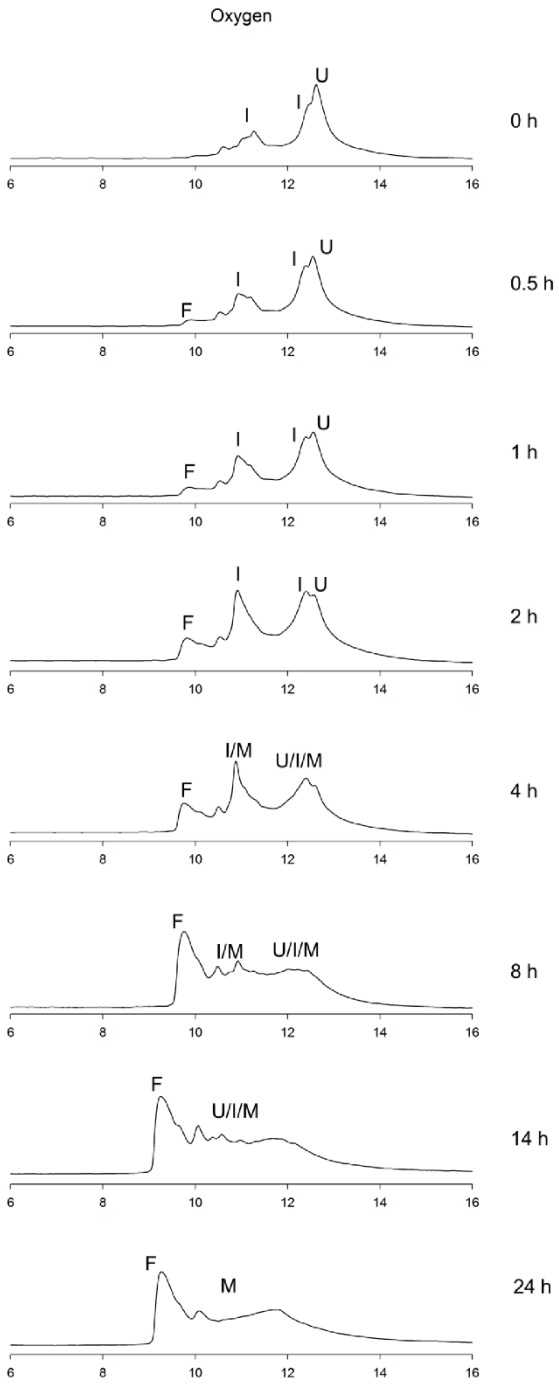


Figure S6. HPLC monitoring of **2** in ammonium acetate (mixture A) at 220 nm (*10→60*).

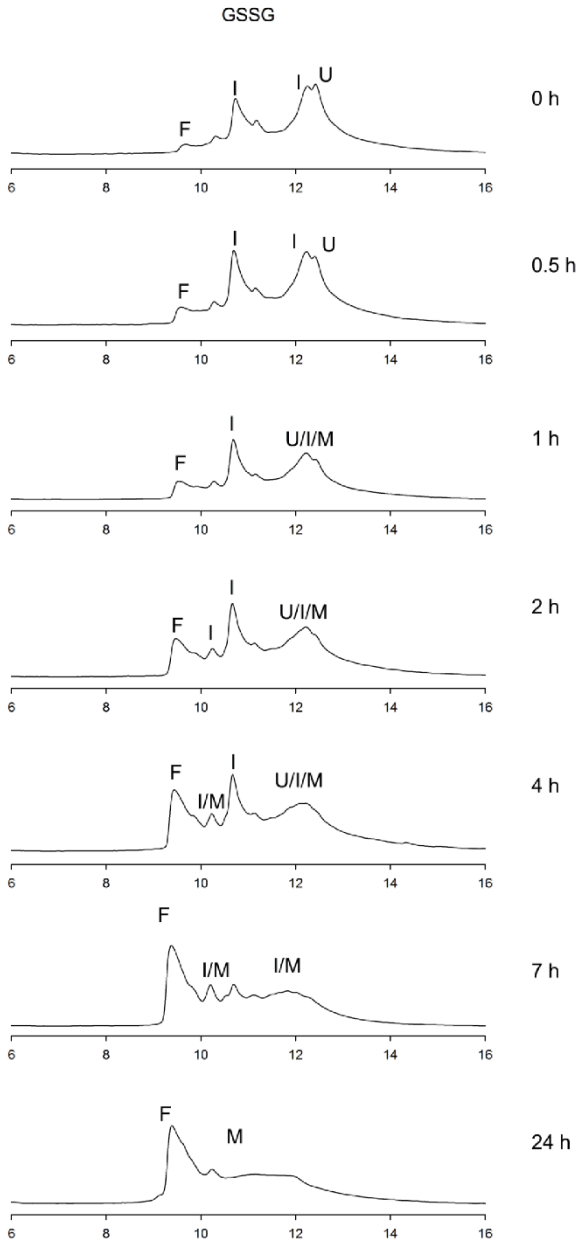


Figure S7. HPLC monitoring of **2** in aqueous ox/red.glutathione (mixture B) at 220 nm (*10→60*).

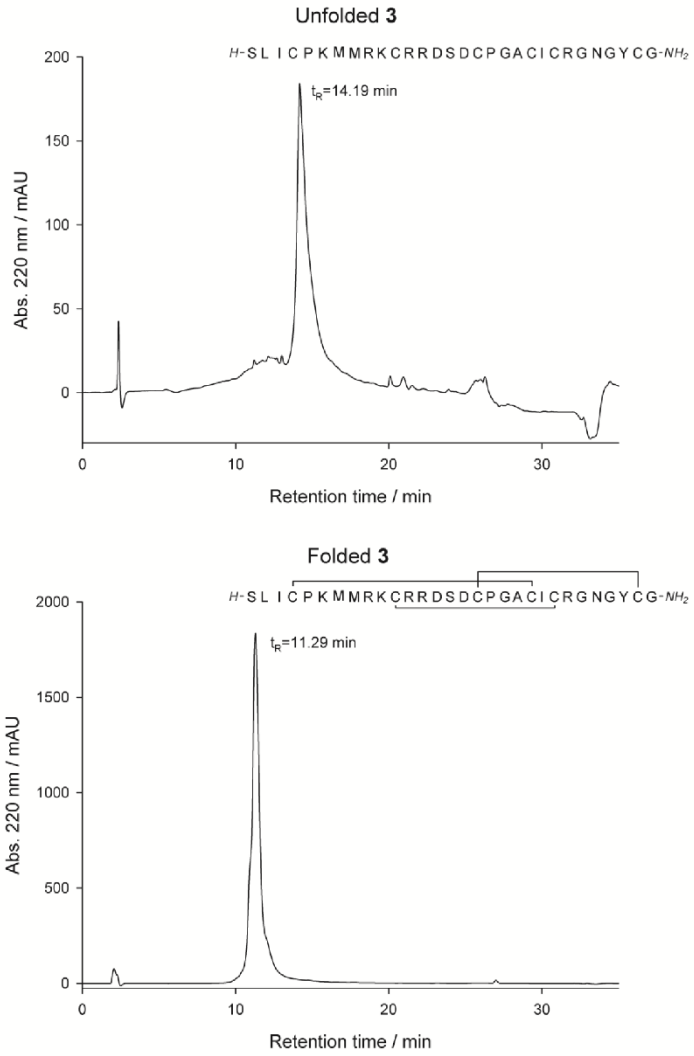


Figure S8. HPLC trace at 220 nm (10→60) of **3** in unfolded and folded state.

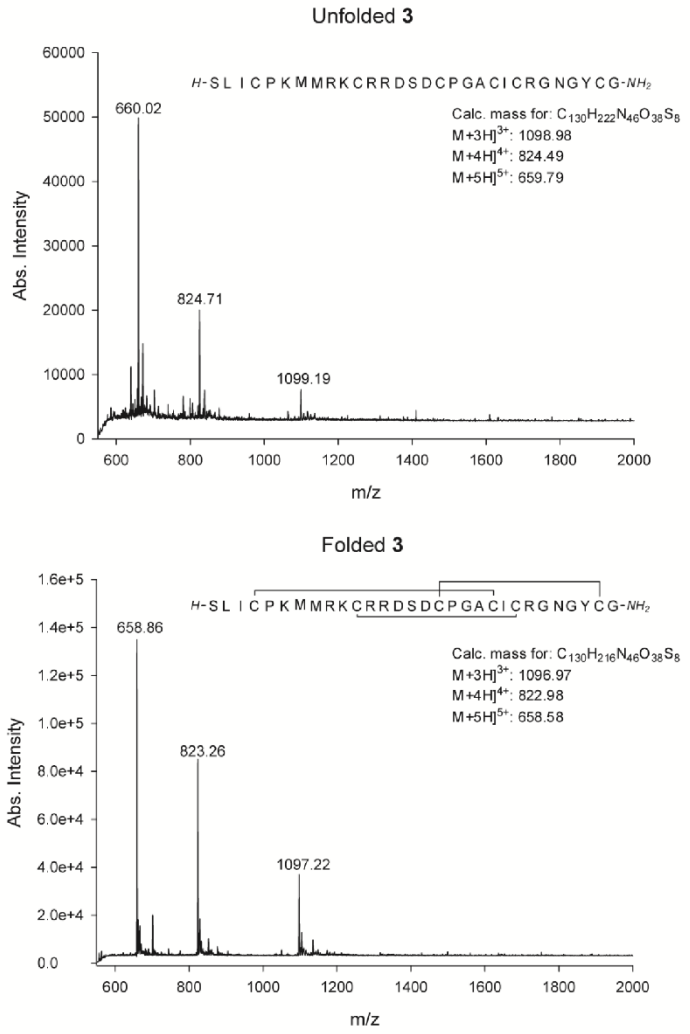


Figure S9. ESI-MS spectra of **3** in unfolded and folded state.

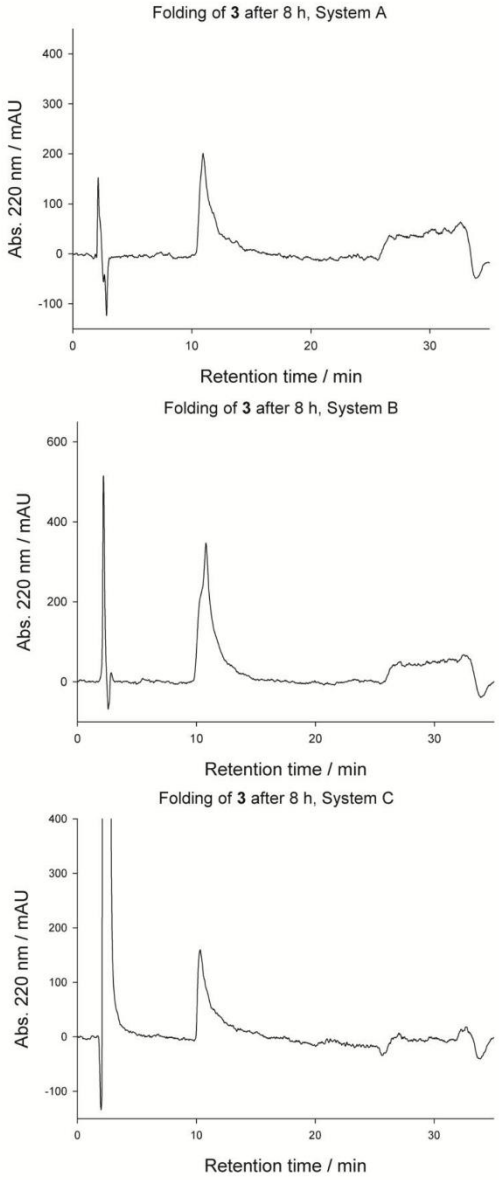


Figure S10. HPLC trace at 220 nm (10→60) of **3** in the corresponding folding mixture after 24 h.

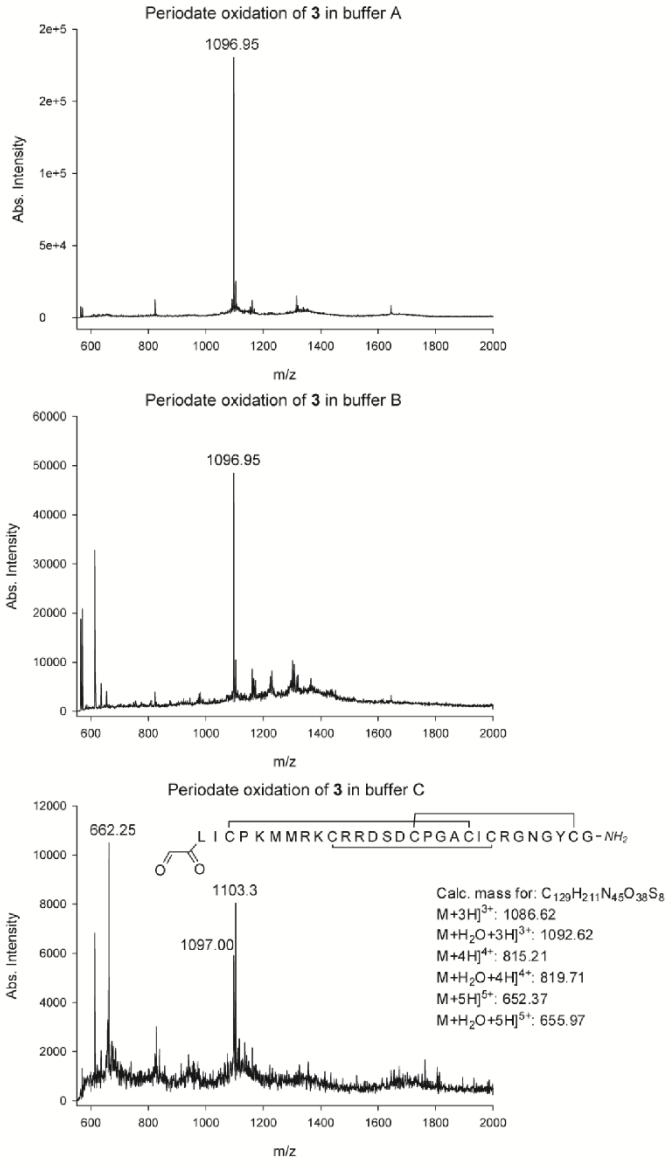


Figure S11. ESI-MS spectra of periodate oxidation of **3** in the folding mixtures after 10 min.

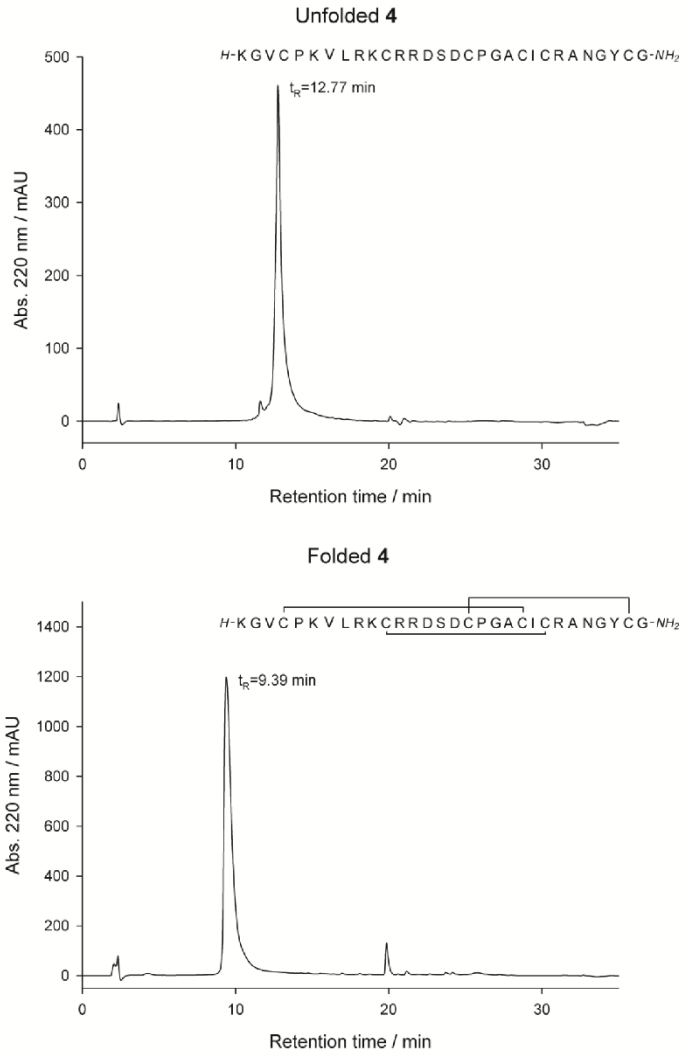


Figure S12. HPLC trace at 220 nm (10→60) of **4** in unfolded and folded state.

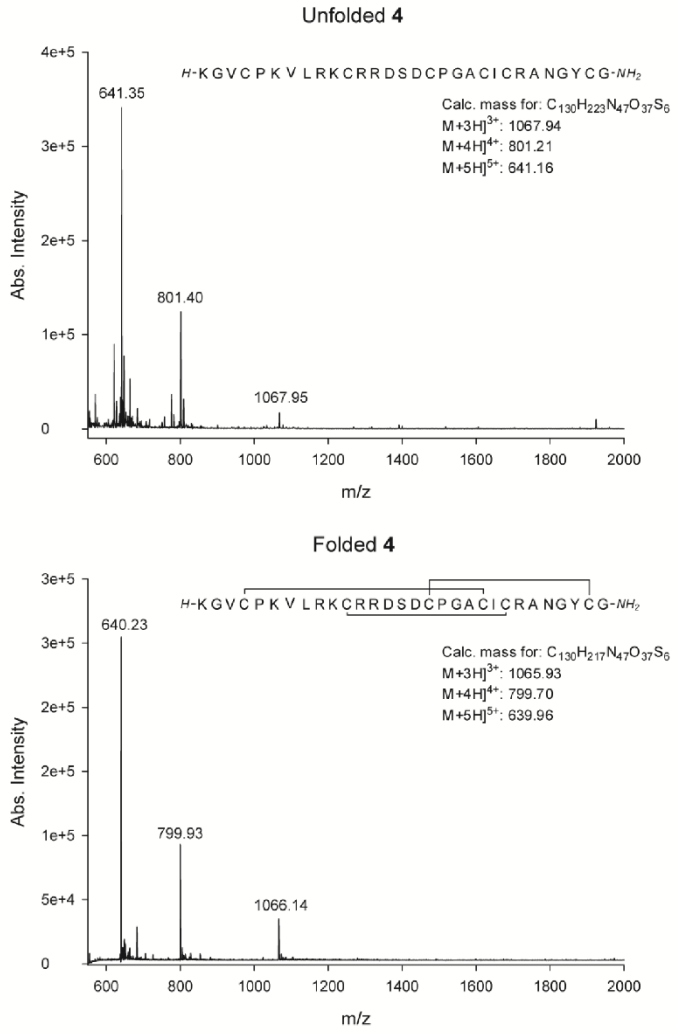


Figure S13. ESI-MS spectra of **4** in unfolded and folded state.

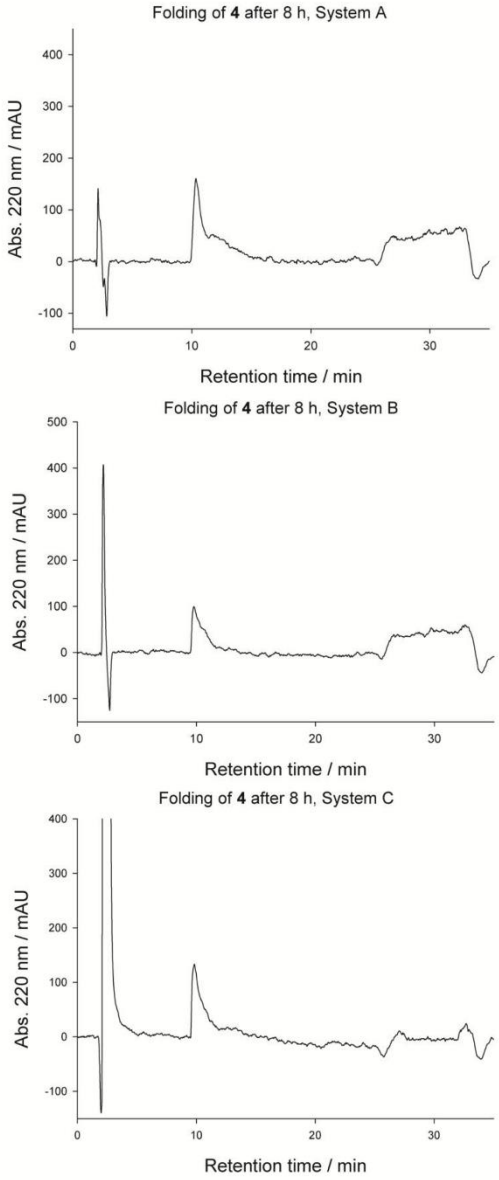


Figure S14. HPLC trace at 220 nm (10→60) of **4** in the corresponding folding mixture after 24 h.

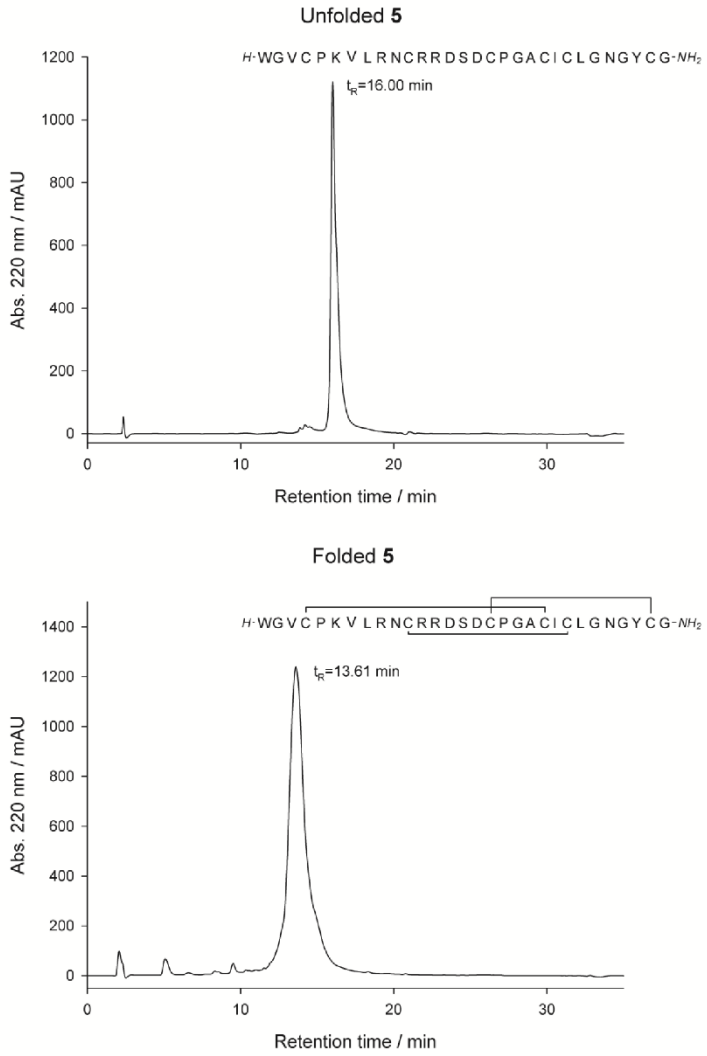


Figure S15. HPLC trace at 220 nm (10→60) of **5** in unfolded and folded state.

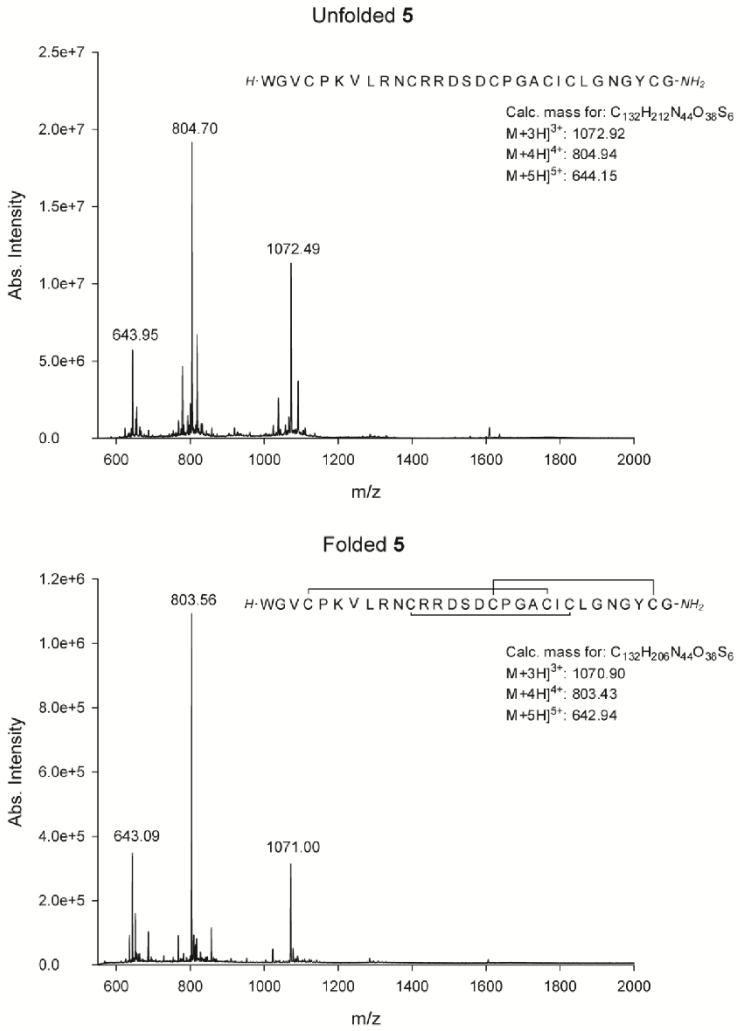


Figure S16. ESI-MS spectra of **5** in unfolded and folded state.

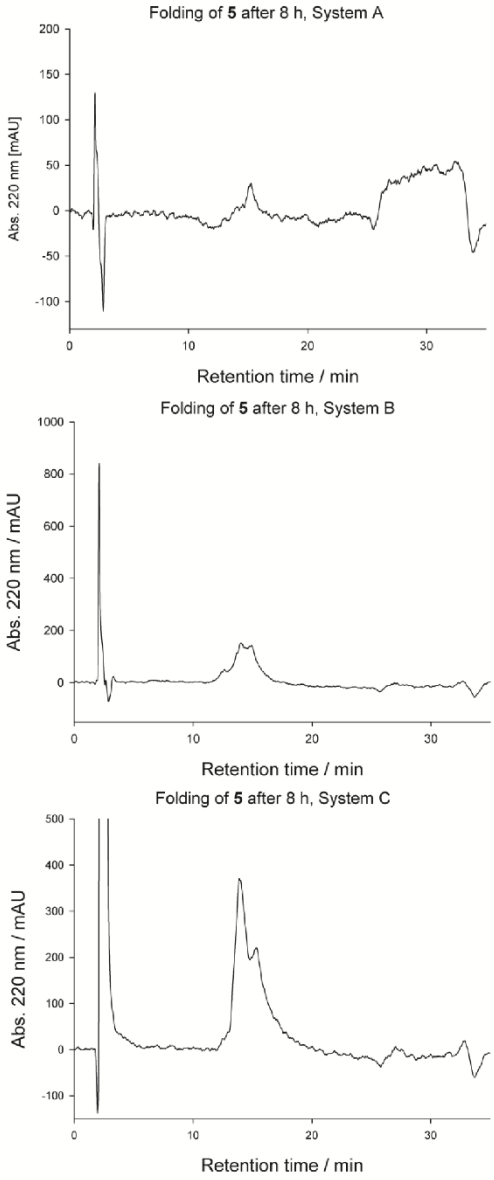


Figure S17. HPLC trace at 220 nm (10→60) of **5** in the corresponding folding mixture after 24 h.

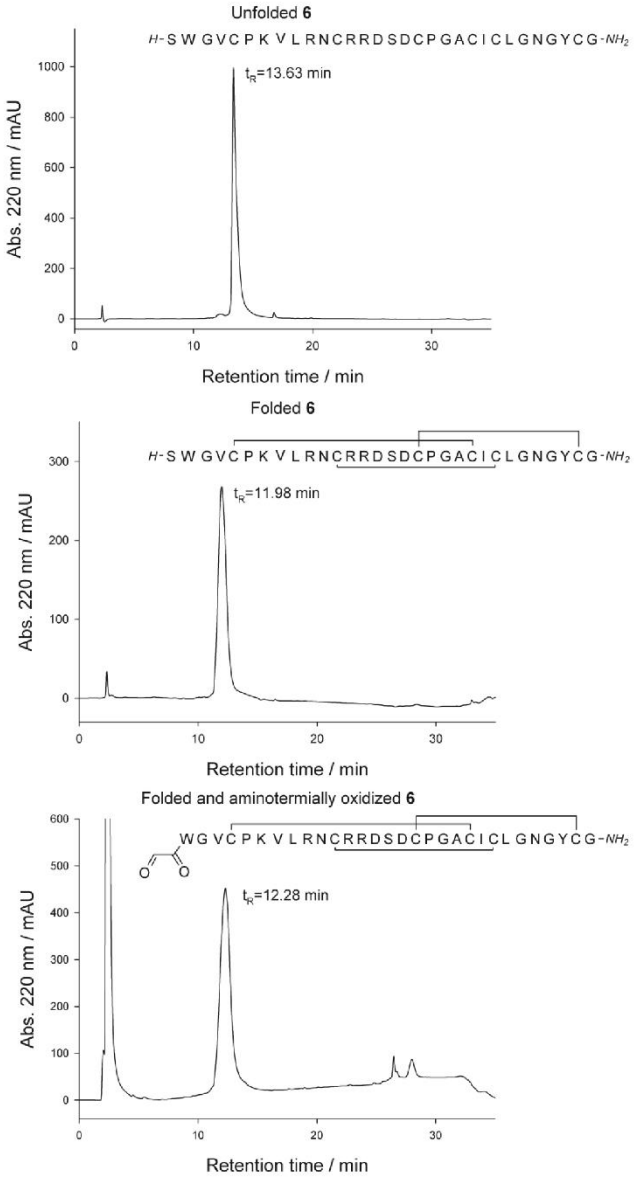


Figure S18. HPLC trace at 220 nm (10→60) of **6** in unfolded, folded and folded and aldehydic state.

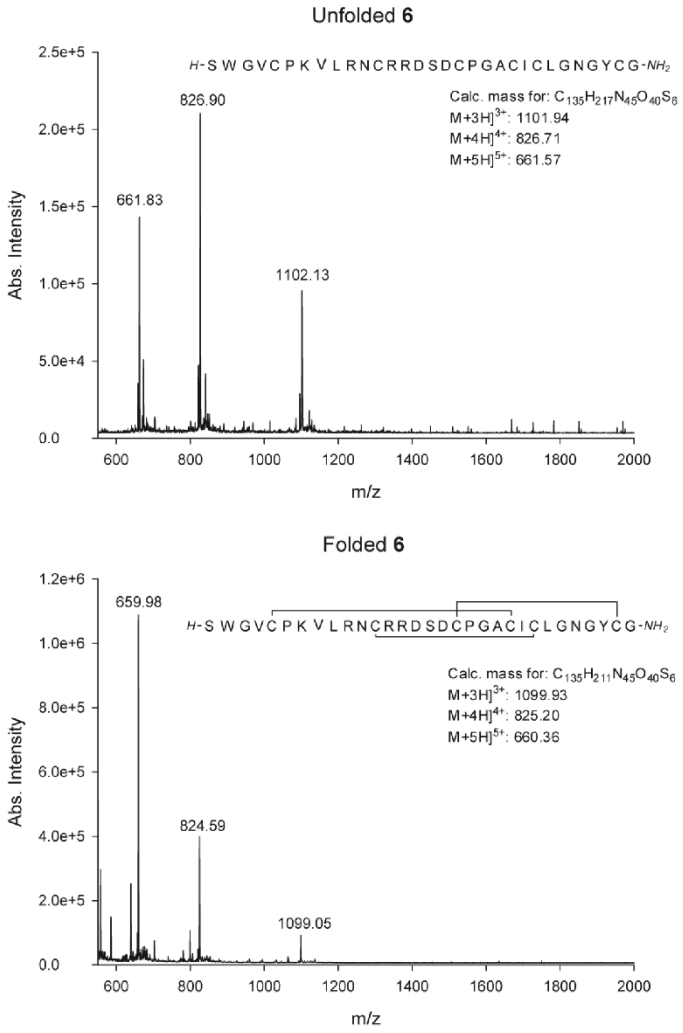


Figure S19. ESI-MS spectra of **6** in unfolded and folded state.

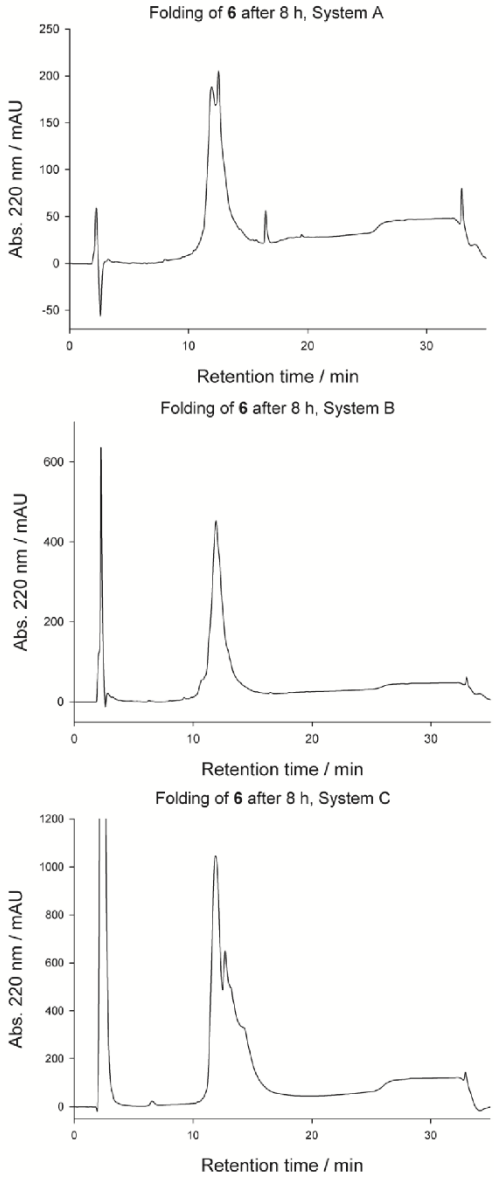


Figure S20. HPLC trace at 220 nm (10→60) of **6** in the corresponding folding mixture after 24 h.

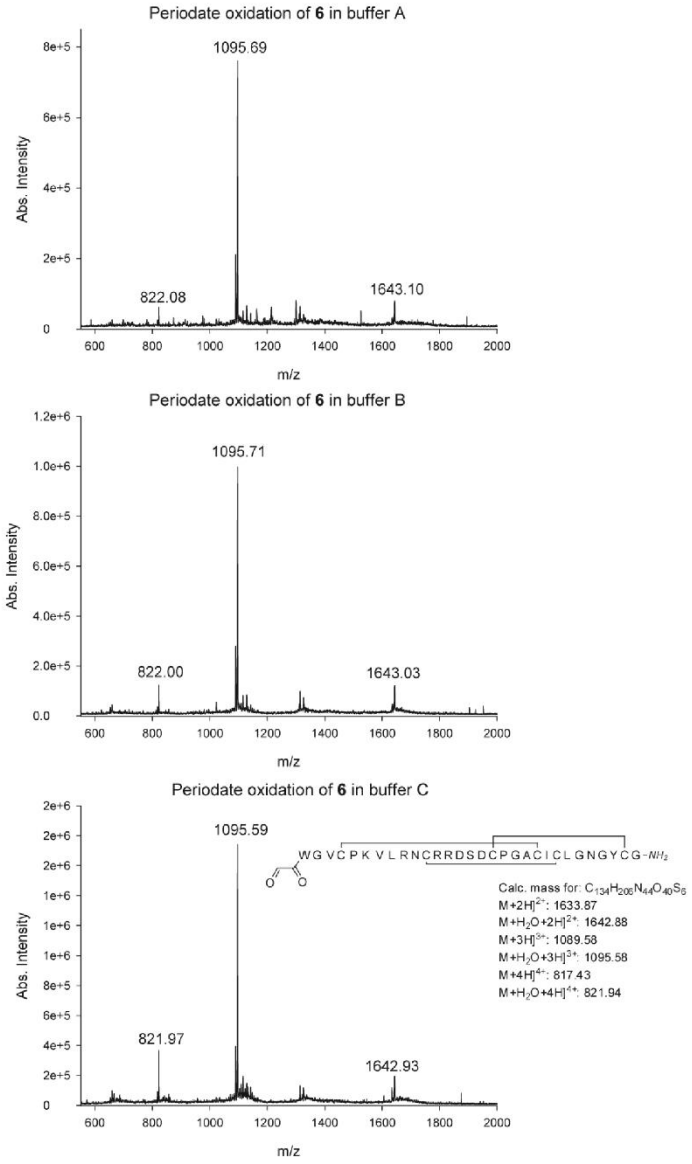


Figure S21. ESI-MS spectra of periodate oxidation of **6** in the folding mixtures after 10 min.

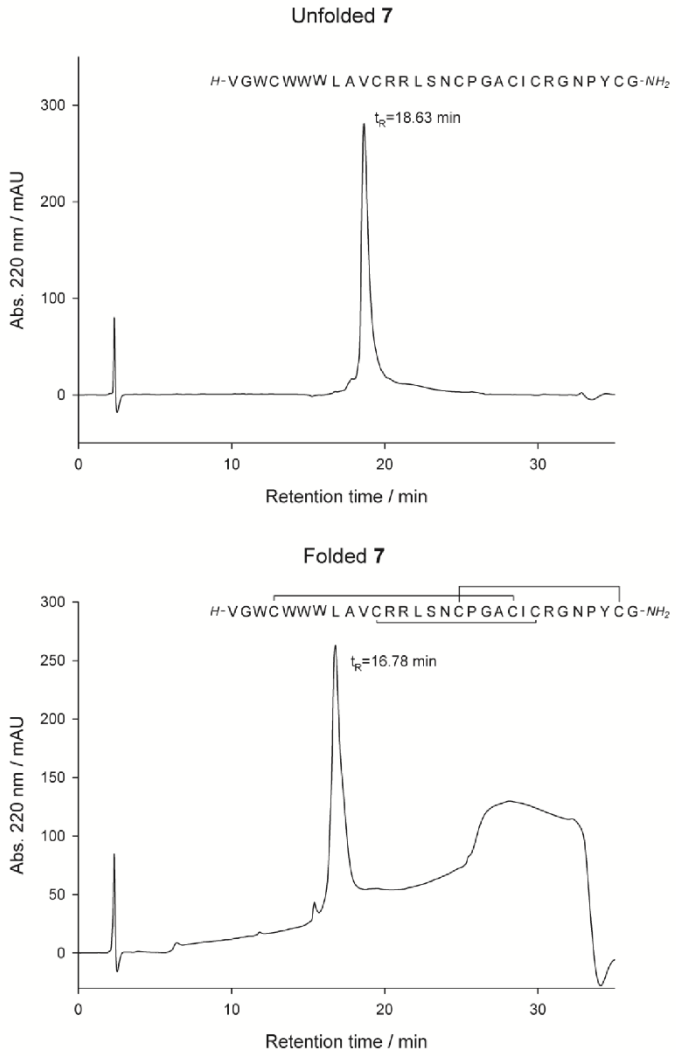


Figure S22. HPLC trace at 220 nm (10→60) of **7** in unfolded and folded state.

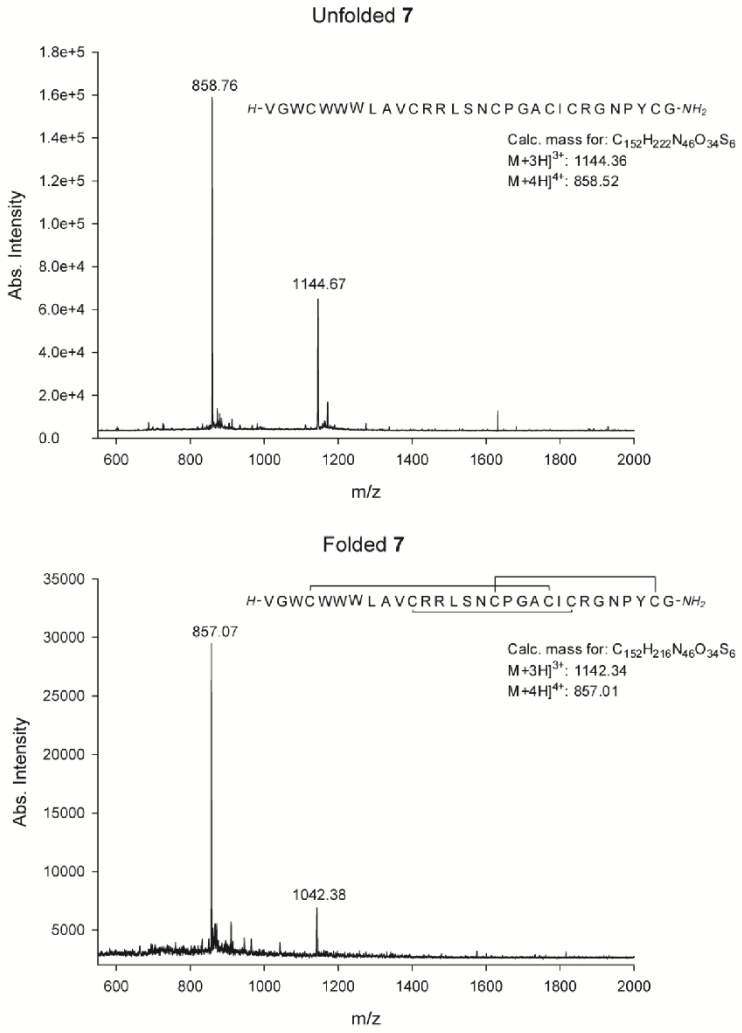


Figure S23. ESI-MS spectra of **7** in unfolded and folded state.

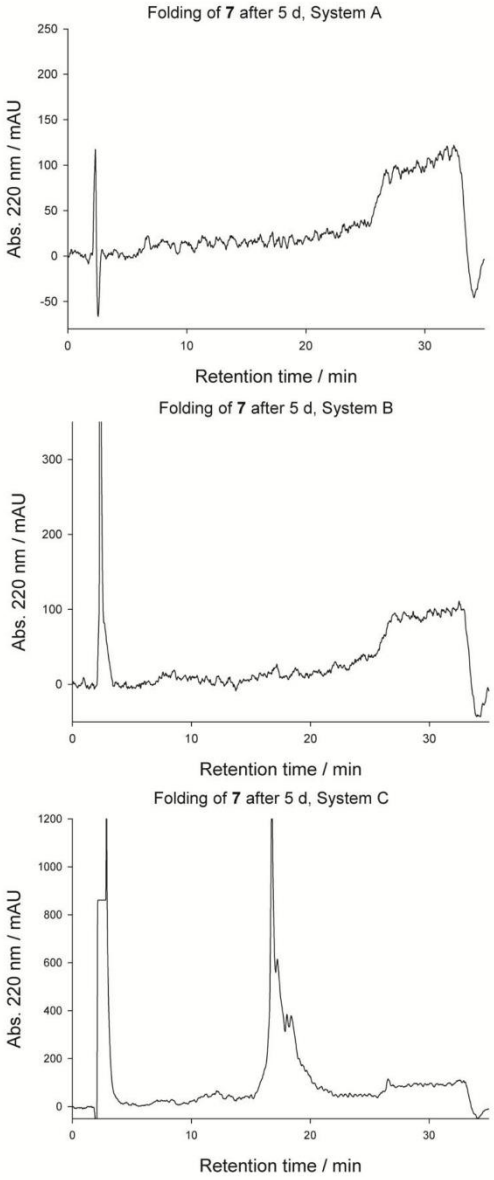


Figure S24. HPLC trace at 220 nm (10→60) of **7** in the corresponding folding mixture after 24 h.

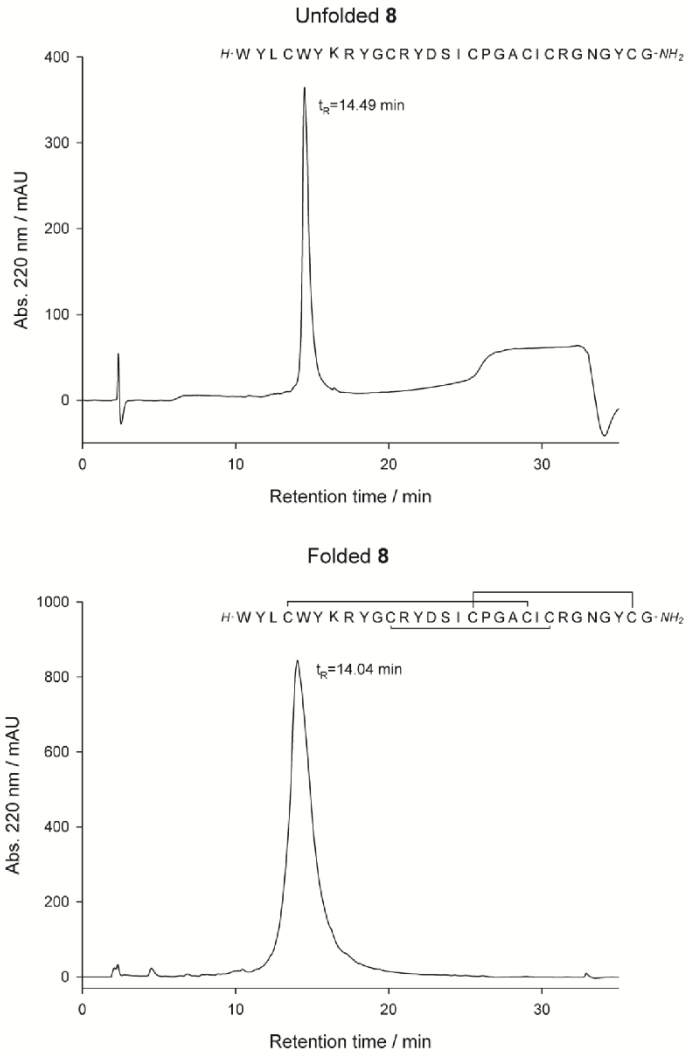


Figure S25. HPLC trace at 220 nm (10→60) of **8** in unfolded and folded state.

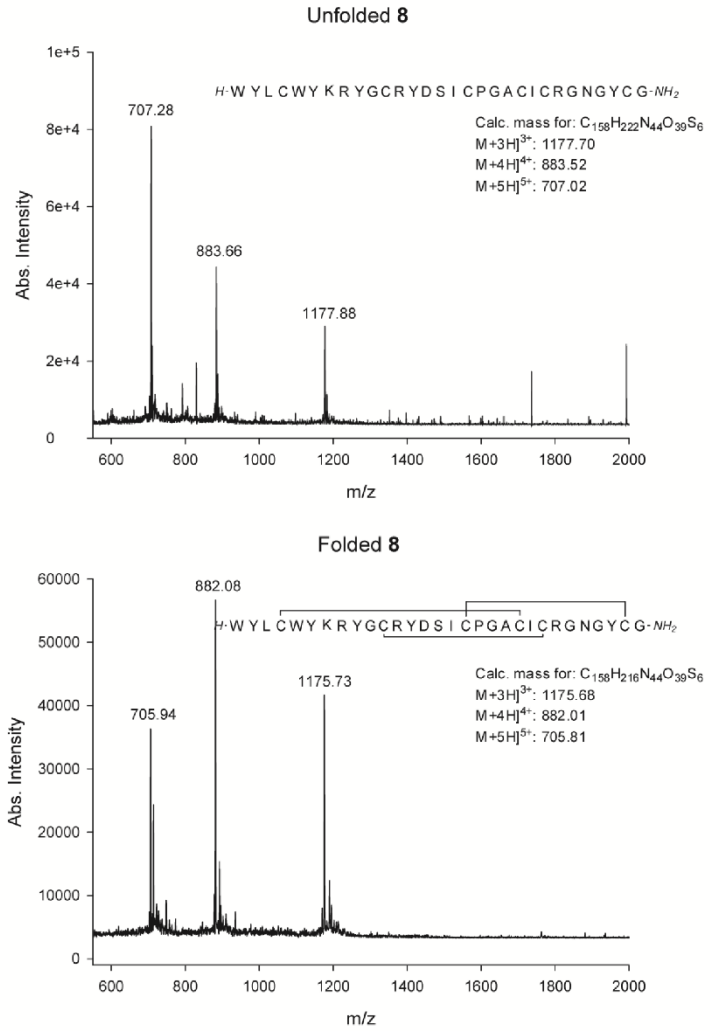


Figure S26. ESI-MS spectra of **8** in unfolded and folded state.

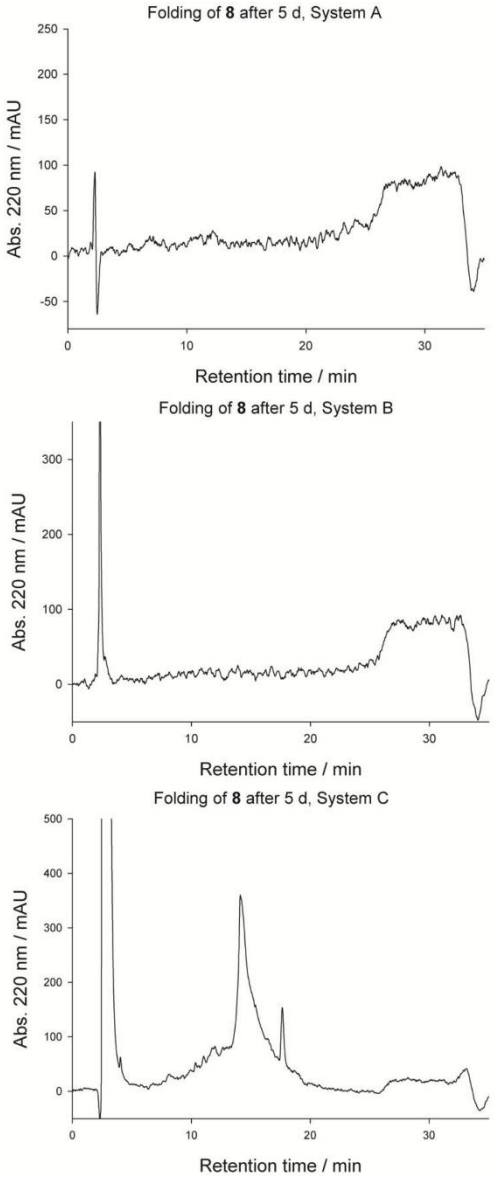


Figure S27. HPLC trace at 220 nm (10→60) of **8** in the corresponding folding mixture after 24 h.

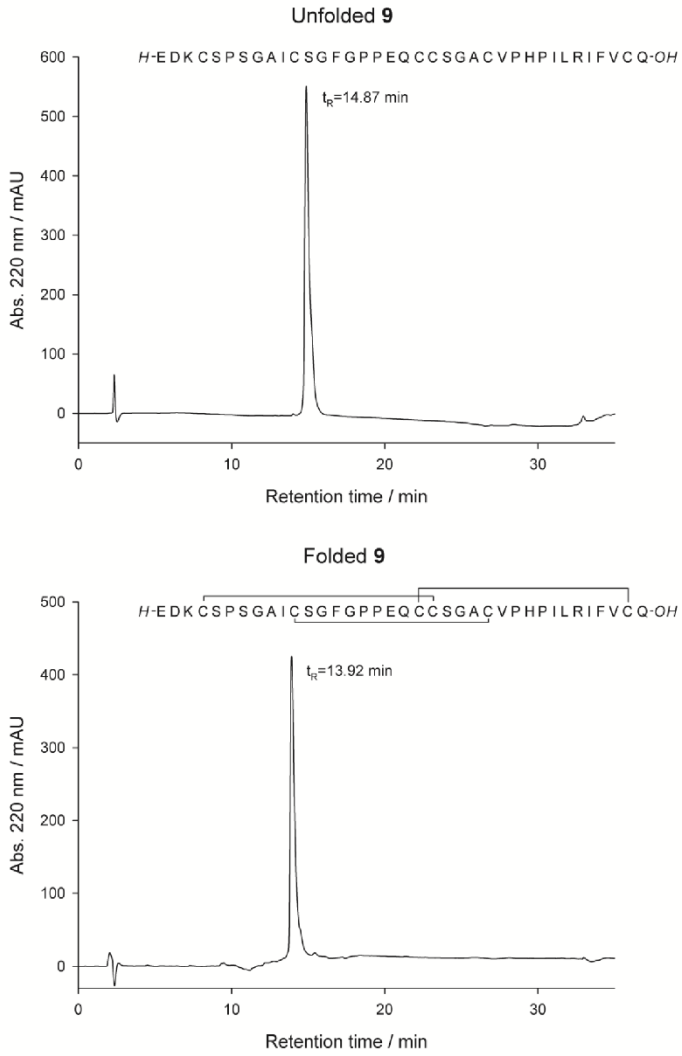


Figure S28. HPLC trace at 220 nm (10→80) of **9** in unfolded and folded state.

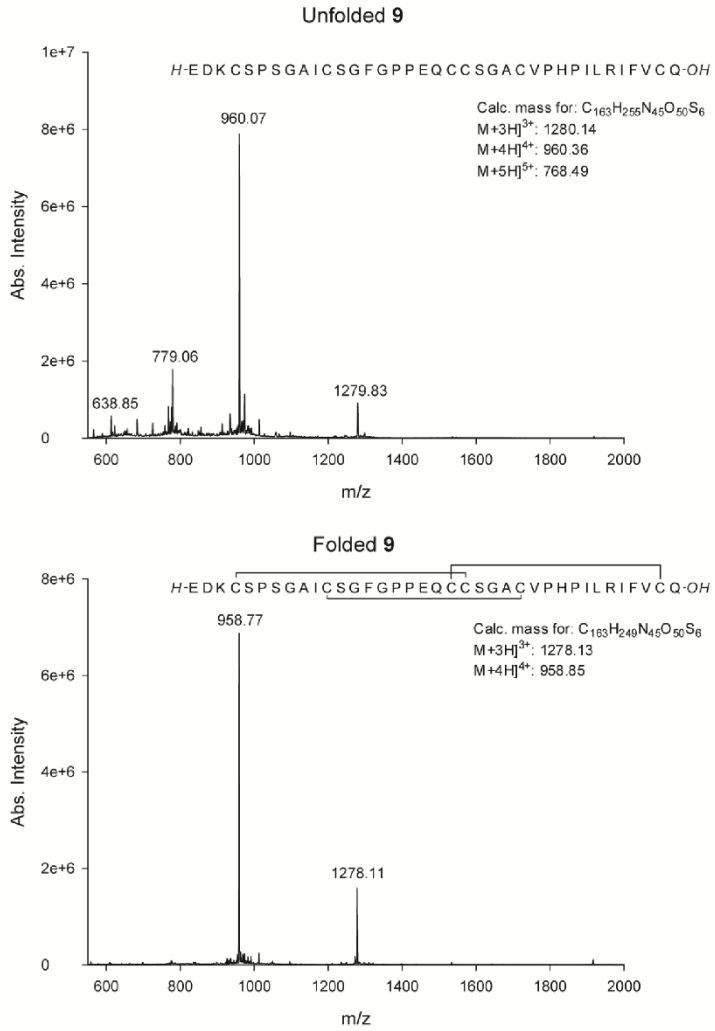


Figure S29. ESI-MS spectra of **9** in unfolded and folded state.

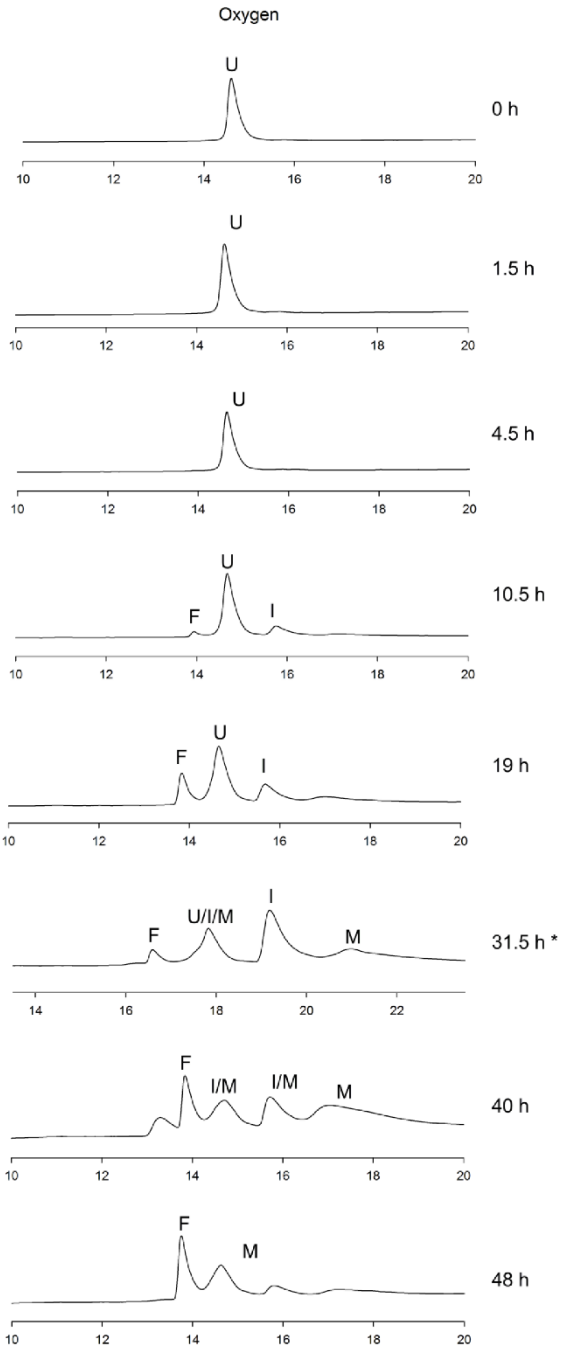


Figure S30. HPLC monitoring of **9** in ammonium acetate (mixture A) at 220 nm (*10→80*). * indicates a different HPLC gradient.

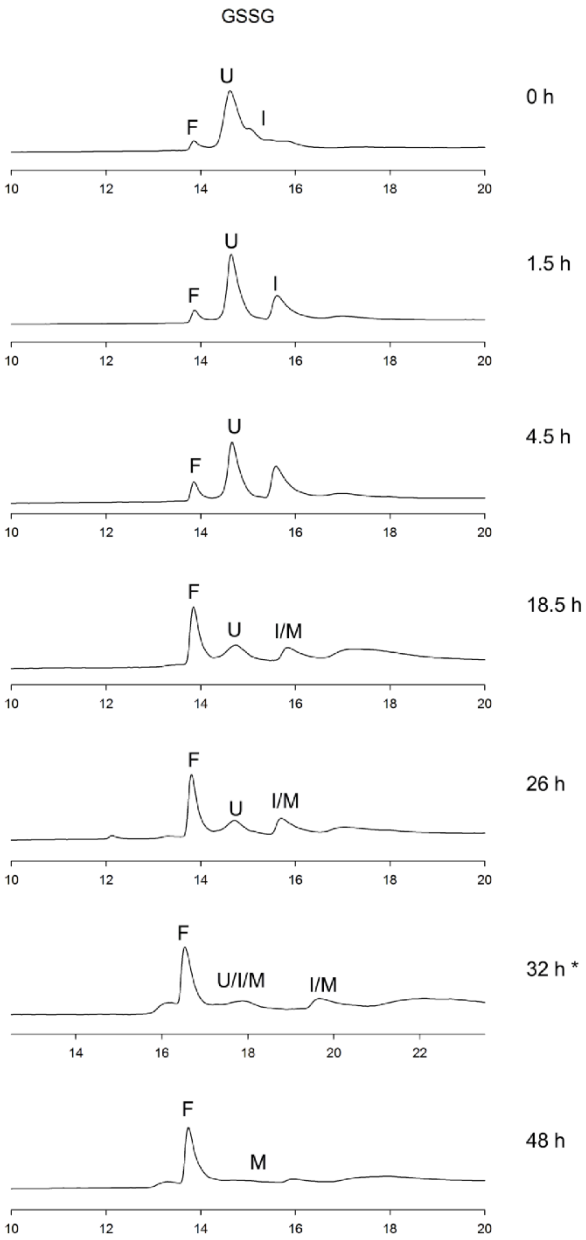


Figure S31. HPLC monitoring of **9** in aqueous ox/red.glutathione (mixture B) at 220 nm (*10→80*). * indicates a different HPLC gradient.

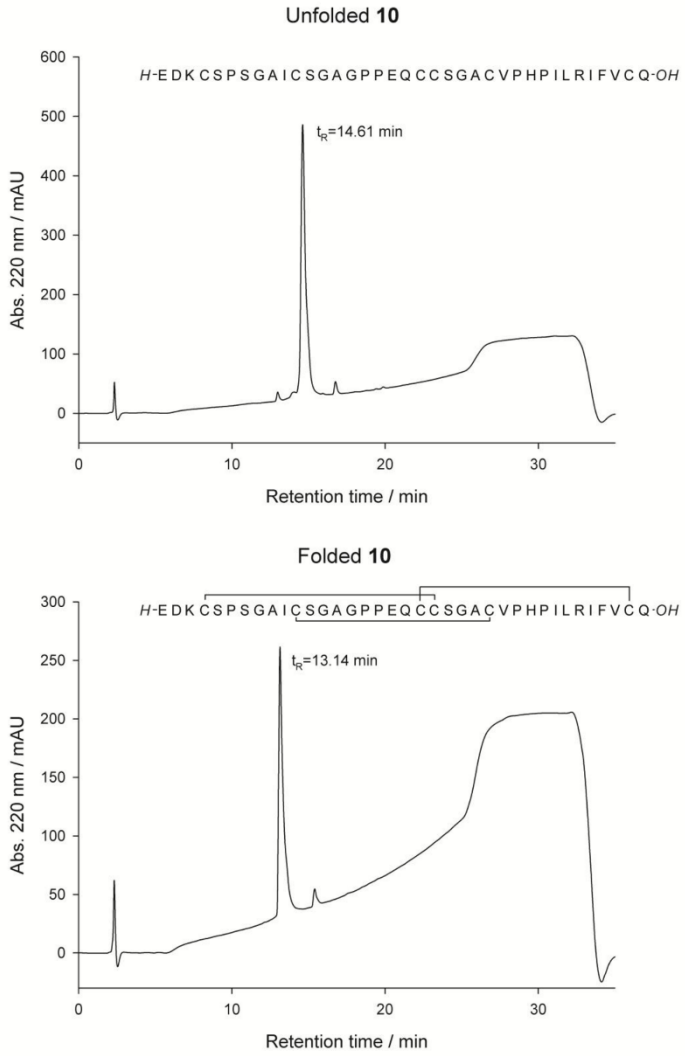


Figure S32. HPLC trace at 220 nm (10→80) of **10** in unfolded and folded state.

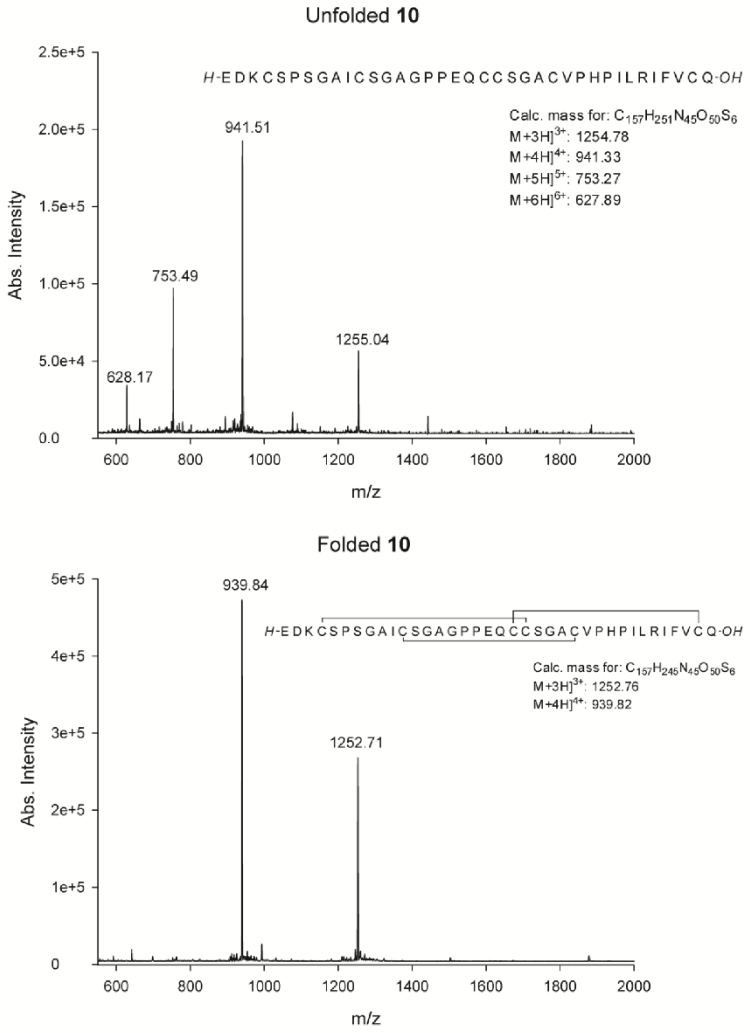


Figure S33. ESI-MS spectra of **10** in unfolded and folded state.

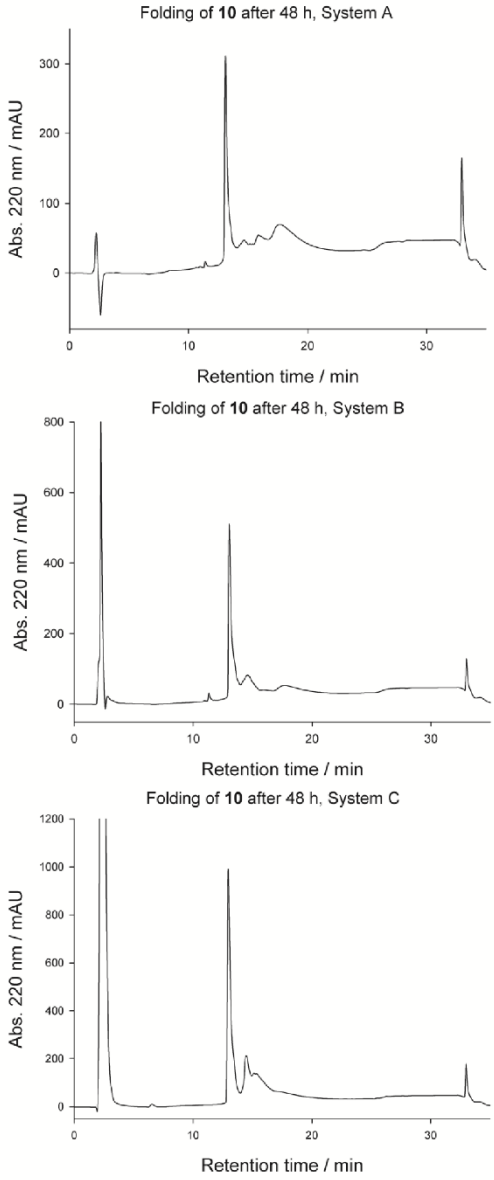


Figure S34. HPLC trace at 220 nm (10→80) of **10** in the corresponding folding mixture after 24 h.

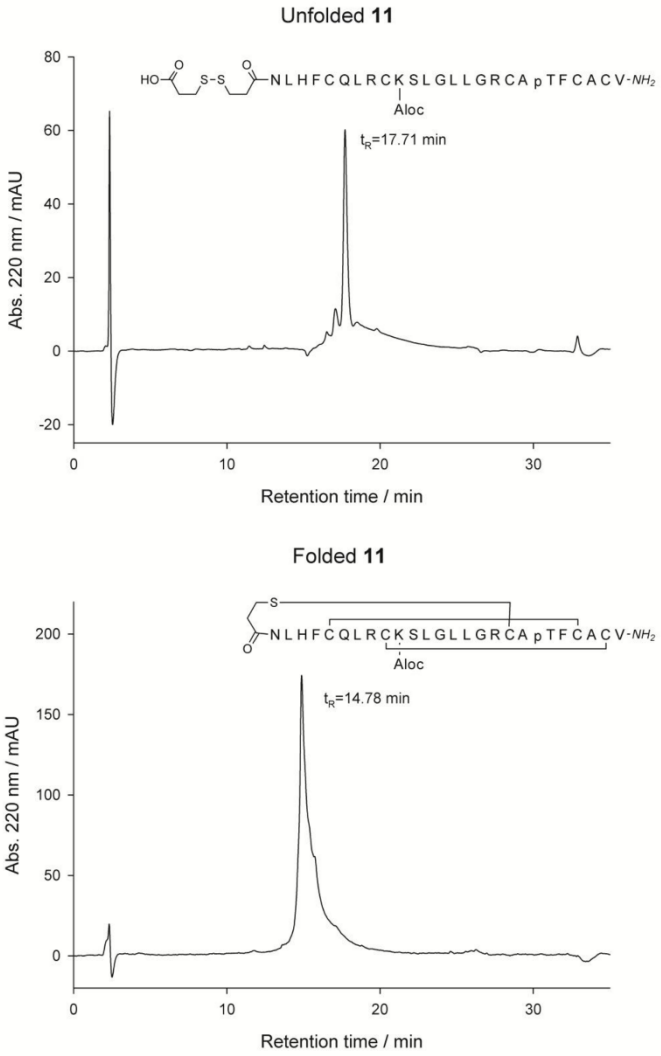


Figure S35. HPLC trace at 220 nm (10→80) of **11** in unfolded and folded state.

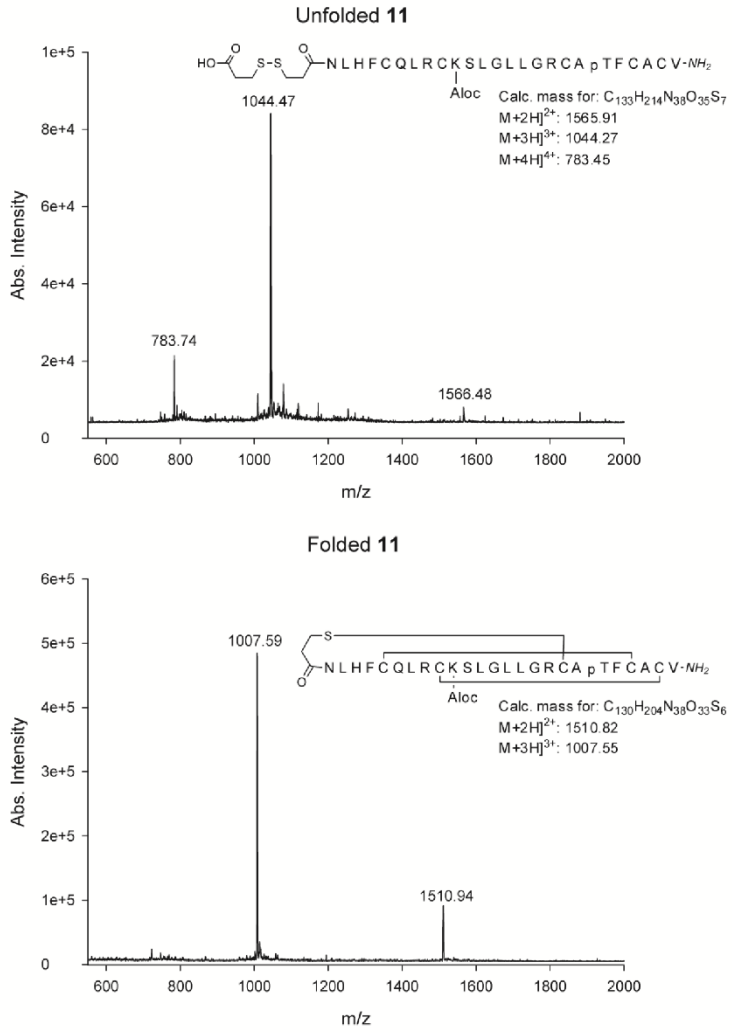


Figure S36. ESI-MS spectra of **11** in unfolded and folded state.

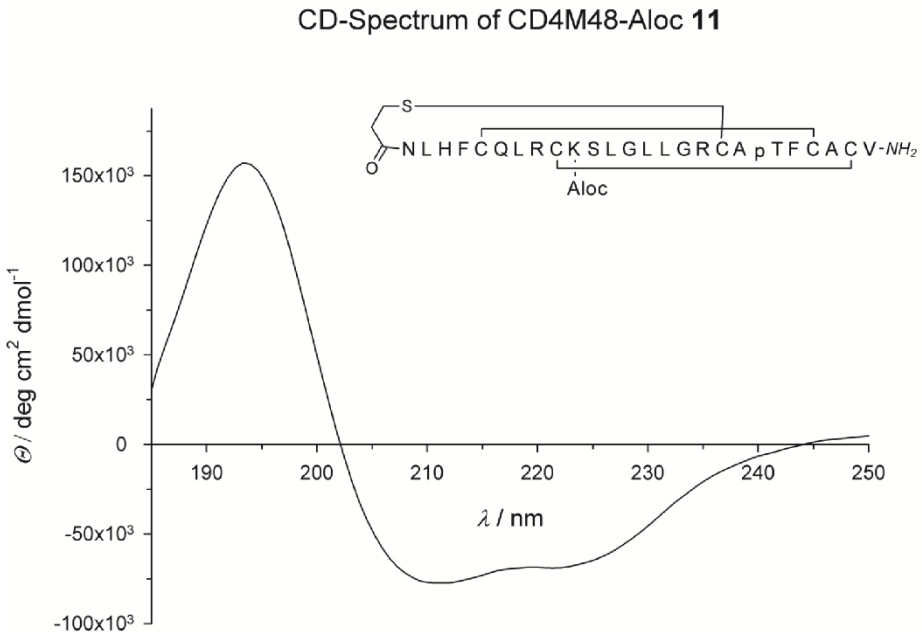


Figure S37. CD spectrum of **11** in folded state in 2 mM aq. Na_2HPO_4. Accumulation of 100 single determinations including background subtraction of 2 mM aq. Na_2HPO_4. Smoothed with the 'smooth function' of Sigma Plot 11.

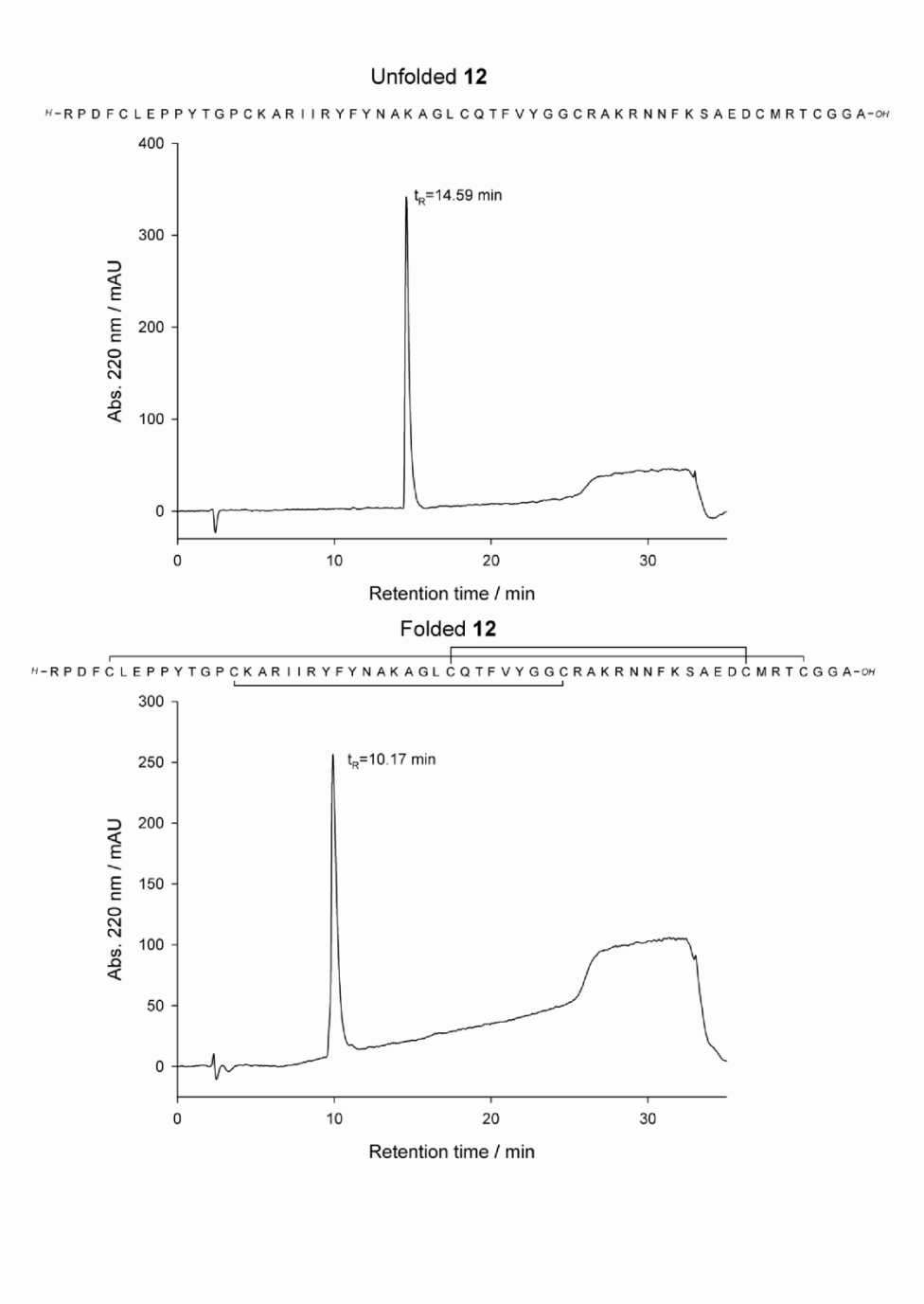


Figure S38. HPLC trace at 220 nm (10→80) of **12** in unfolded and folded state.

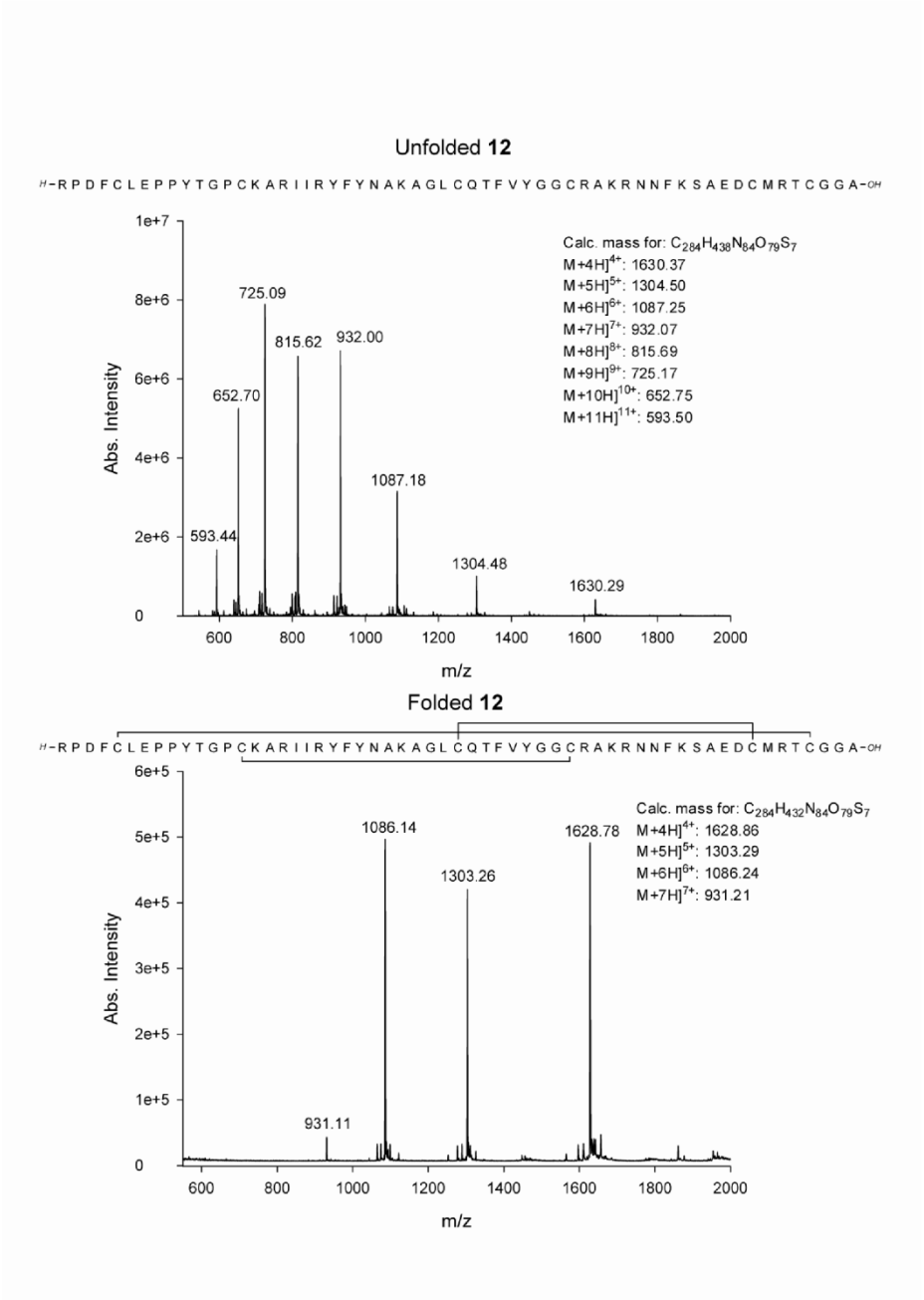


Figure S39. ESI-MS spectra of **12** in unfolded and folded state.

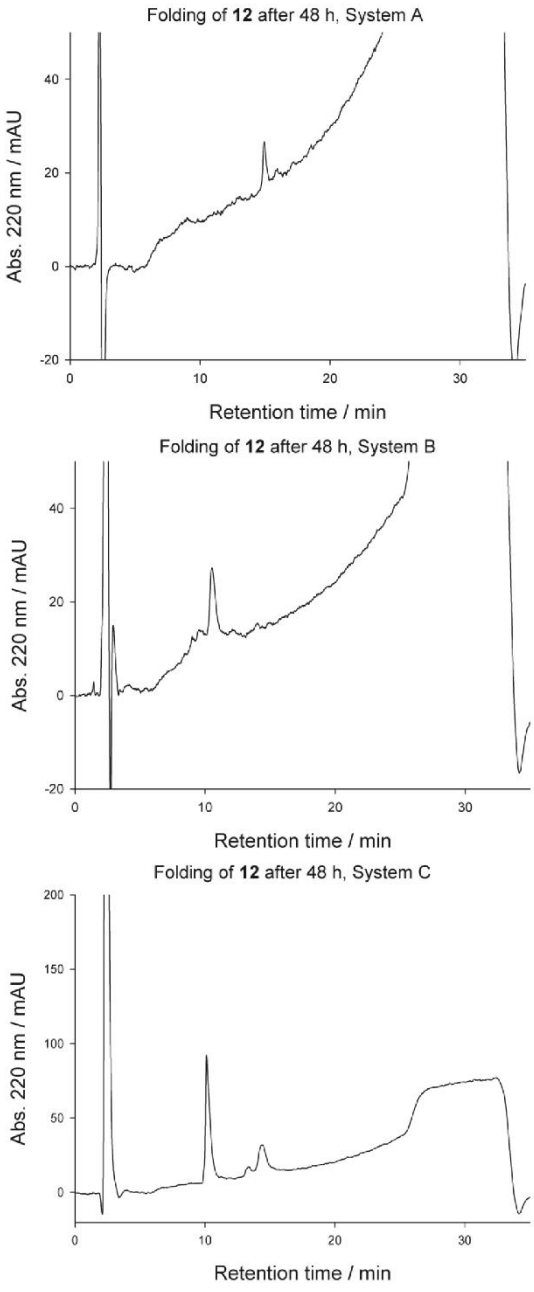


Figure S40. HPLC trace at 220 nm (10→60) of **12** in the corresponding folding mixture (in the case of System C the reaction mixture was tenfold diluted before injection into the HPLC system) after 48 h.

Supporting information

Combinatorial optimization of cystine-knot peptides towards high-affinity inhibitors of human matriptase-1

Potent cystine-knot inhibitors of matriptase-1

Bernhard Glotzbach[1.], Michael Reinwarth[1.], Niklas Weber[1], Sebastian Fabritz[2], Michael Tomaszowski[1], Heiko Fittler[1], Andreas Christmann[1], Olga Avrutina[1], and Harald Kolmar*[1]

1 Clemens-Schöpf-Institut für organische Chemie und Biochemie, Technische Universität Darmstadt,

2 AB SCIEX Germany GmbH

* Corresponding author:

Tel: (+) 49 6151 164742

Fax:(+)49 6151 165399

E-mail: Kolmar@Biochemie-TUD.de

. These authors contributed equally to this work

Table of Contents:

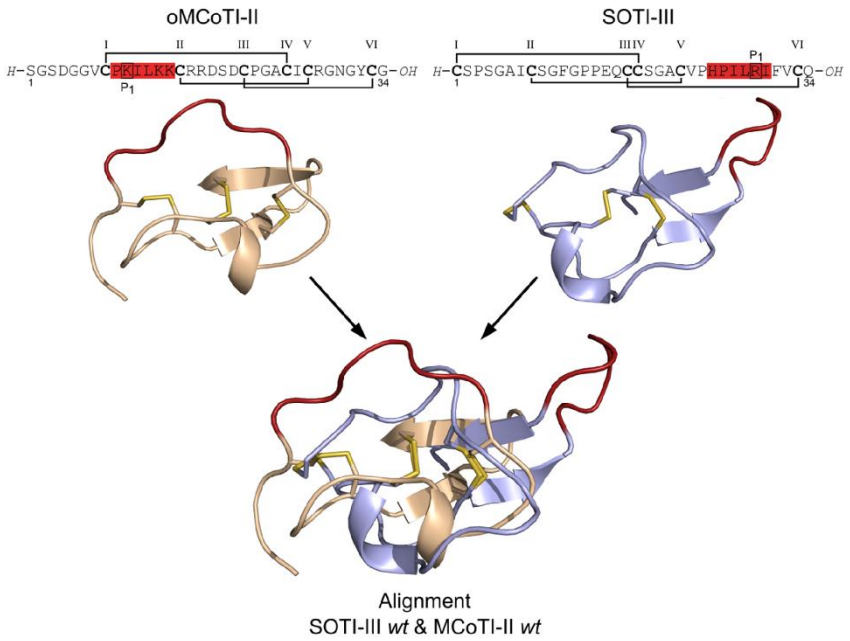


Figure S1: Sequences and structure alignment of cystine-knot trypsin inhibitors oMCoTI-II (light brown, pdb: 1ha9, upper left)[1] and SOTI-III (light blue, pdb: 4aor, upper right).[2] Secondary structure is shown as cartoon and cysteine residues are depicted as yellow sticks; protease-binding regions are depicted in red. Cystine-forming residues are marked bold, and the numbering of respective cysteines is according to their appearance in the sequence

Isolation cyclic MCoTI variants from the seeds of *Momordica cochinchinensis*

Cyclic miniproteins were isolated as previously reported from 5 g of homogenized seeds. Extraction was performed using 20 mL aqueous sodium acetate (20 mM, pH 4.5) at ambient temperature for 16 h.[3] The suspension was filtrated and proteins were denatured with 20 mL aqueous acetone (40 %, v/v), while the miniproteins remained in their native conformation. After removal of acetone under reduced pressure, the suspension was filtered and the filtrate was purified by semi-preparative HPLC using an axia-packed Phenomenex Luna C18 (250 x 21.2 mm, 5 µm, 100 Å) column applying linear acetonitrile gradients at a flow rate of 10 mL/min. Isocratic elution (10 % eluent B over 5 was followed by a linear gradient of 10→55 % B over 30 min (Figure S2).

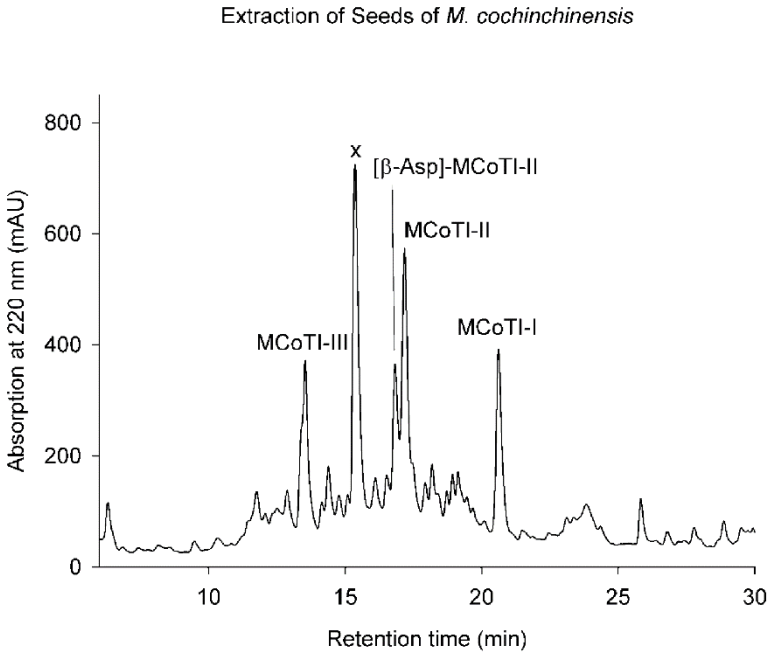
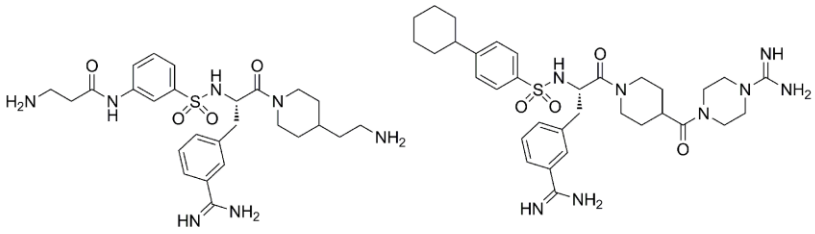


Figure S2: HPLC trace of MCoTI-variants isolated out of seeds of *Momordica cochinchinensis*. x marks an unidentified peak. [β-Asp]-MCoTI-II possesses a β-aspartyl residue at position 4.[3]

Table S1: Apparent inhibition constants towards matriptase-1 of the isolated cyclic MCoTI variants.

Miniprotein	Amino acid sequence of cyclic MCoTI variants; inhibitor loop between CysI and CysII is marked bold	K_i^{app} (nM)	Yield (from 5 g seeds) (mg)
MCoTI-I	SGSDGGVC**PKILQR**CRRDSDCPGACICRGNGYCG	>1000	1.5
MCoTI-II	SGSDGGVC**PKILKK**CRRDSDCPGACICRGNGYCG	235	1.0
oMCoTI-II[a]	SGVC**PKILKK**CRRDSDCPGACICRGNGYCG	166	n. d.
MCoTI-II (ß-asp)	SGS(β-D)GGVC**PKILKK**CRRDSDCPGACICRGNGYCG[b]	390	1.4
MCoTI-III[a]	(pyro-E)RAC**PRILKK**CRRDSDCPGECICKENGYCG[c]	275	1.9

a oMCoTI-II and MCoTI-III possess a linear backbone,b: (β-D) stands for a β-aspartyl bond, c: pyro-E stands for pyroglutamate eliminating cyclization.[3]



S1 **S2**

Figure S3: Small-molecule inhibitors of matriptase-1 that were used as reference compounds.[4]

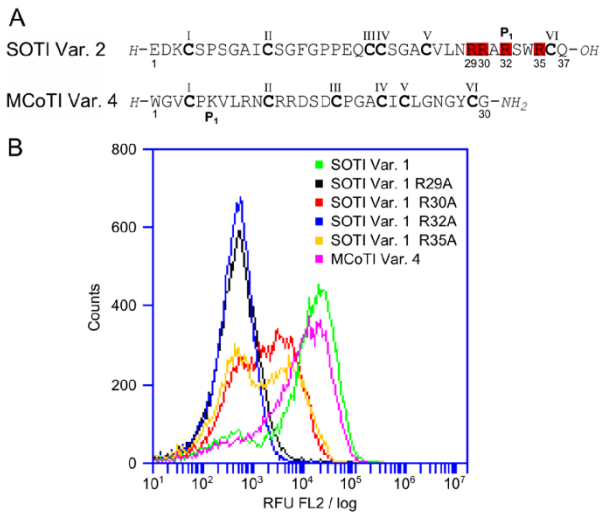


Figure S4: Matriptase-1 binding analysis of miniprotein variants SOTI Var. 1 and MCoTI Var. 4 *via* flow cytometry. A) Sequence of the isolated matriptase-1 inhibitors with randomized residues depicted in the according color. B) Overlay of FACS histograms after labeling of miniprotein-displaying yeast cells with 1 µM of biotinylated matriptase-1 followed by incubation with Streptavidin, R-phycoerythrin conjugate.

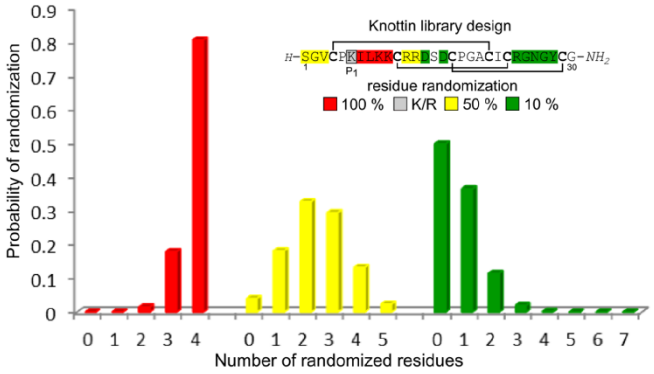


Figure S5: Expected distribution of the appearance of amino acid exchanges in loop 1 (red), flanking regions of loop 1 (yellow), and loop 4 (green). The calculation was performed assuming a binominal distribution function.

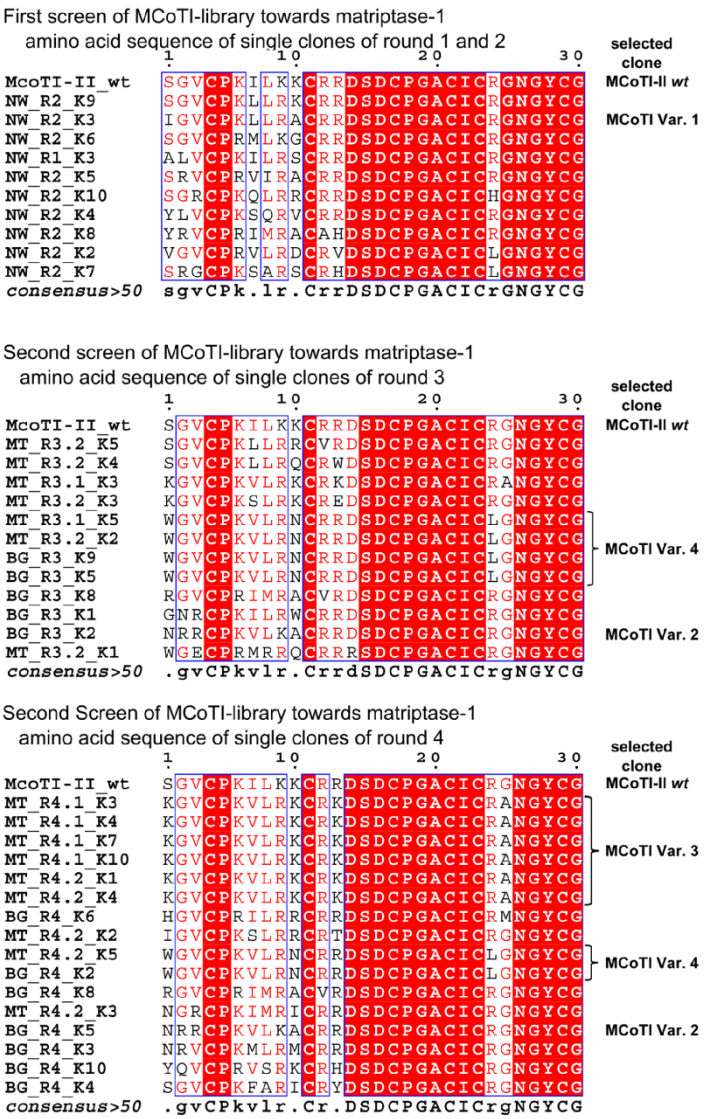


Figure S6: Sequence alignments of MCoTI variants isolated from two screening cycles. Multiple sequence alignments were performed with MultAlin.[5] Amino acids marked in red are identical to those of the MCoTI-*wt***3**; amino acids highlighted in red are conserved for all aligned sequences. The blue frames show the consensus of at least two amino acids. The consensus sequence (bottom line) was calculated with a threshold of 0.5. Consensus sequence: upper-case letters indicate sequential identity, lower-case letters illustrate consensus. MCoTI *wt* was taken as lead sequence for the alignment. Sequences that were selected for chemical peptide synthesis and further studies are marked on the right.

Table S2: Characterization of synthetic miniproteins.

Miniprotein	HPLC Retention time (min)	Calculated masses	Measured masses	Yield (pure, according to crude product)
SOTI *wt* [6]	13.92 [a]	$M+3H]^{3+}$: 1278.13	$M+3H]^{3+}$: 1278.11	10.8 mg
		$M+4H]^{4+}$: 958.85	$M+4H]^{4+}$: 958.77	23.8 %
SOTI Var. 1	12.78 [a]	$M+3H]^{3+}$: 1319.14	$M+3H]^{3+}$: 1319.11	4.5 mg
		$M+4H]^{4+}$: 989.61	$M+4H]^{4+}$: 989.54	12.5 %
		$M+5H]^{5+}$: 791.88	$M+5H]^{5+}$: 791.83	
MCoTI *wt* [6]	9.68 [b]	$M+3H]^{3+}$: 1081.26	$M+3H]^{3+}$: 1081.40	8.5 mg
		$M+4H]^{4+}$: 811.19	$M+4H]^{4+}$: 811.45	22.0 %
		$M+5H]^{5+}$: 649.15	$M+5H]^{5+}$: 649.30	
MCoTI Var. 1	12.40 [b]	$M+2H]^{2+}$: 1576.35	$M+2H]^{2+}$: 1576.20	6.2 mg
		$M+3H]^{3+}$: 1051.23	$M+3H]^{3+}$: 1051.08	20.0 %
		$M+4H]^{4+}$: 788.67	$M+4H]^{4+}$: 788.56	
		$M+5H]^{5+}$: 631.14	$M+5H]^{5+}$: 631.06	
		$M+6H]^{6+}$: 526.12	$M+6H]^{6+}$: 526.09	
MCoTI Var. 2 [6]	9.49 [b]	$M+3H]^{3+}$: 1089.59*	$M+3H]^{3+}$: 1089.86	6.4 mg
		$M+4H]^{4+}$: 817.45	$M+4H]^{4+}$: 817.74	19.1 %
		$M+5H]^{5+}$: 654.16	$M+5H]^{5+}$: 654.45	
MCoTI Var. 3 [6]	9.39 [b]	$M+3H]^{3+}$: 1065.93	$M+3H]^{3+}$: 1066.14	7.2 mg
		$M+4H]^{4+}$: 799.70	$M+4H]^{4+}$: 799.93	18.2 %
		$M+5H]^{5+}$: 639.96	$M+5H]^{5+}$: 640.23	
MCoTI Var. 4 [6]	13.61 [b]	$M+3H]^{3+}$: 1070.90	$M+3H]^{3+}$: 1071.00	5.8 mg
		$M+4H]^{4+}$: 803.43	$M+4H]^{4+}$: 803.56	14.1 %
		$M+5H]^{5+}$: 642.94	$M+5H]^{5+}$: 643.09	

a Linear gradient of eluent B from 10 to 80 % over 20 minutes. b Linear gradient of eluent B from 10 to 60 % over 20 minutes.

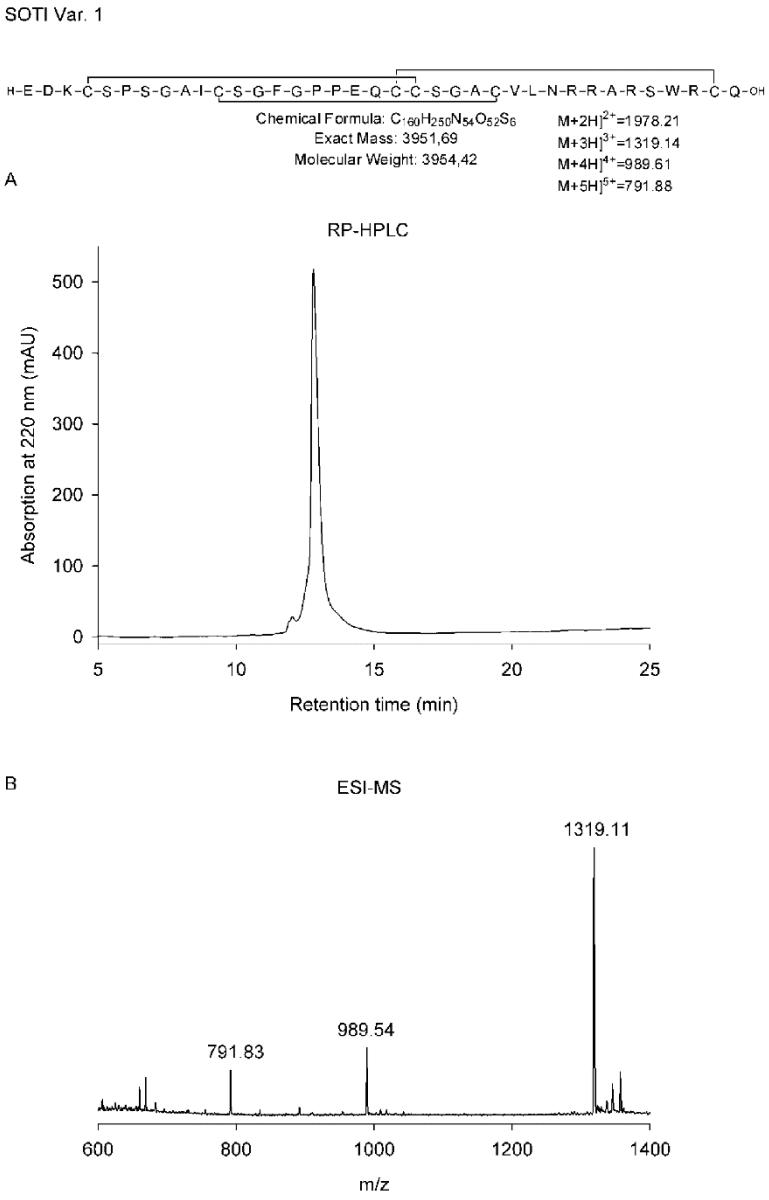


Figure S7: HPLC and MS analysis of folded miniprotein SOTI Var. 1. (**A**) HPLC trace (10 to 80 % B over 20 min) at 220 nm. (**B**) ESI-MS of peptide-containing fraction.

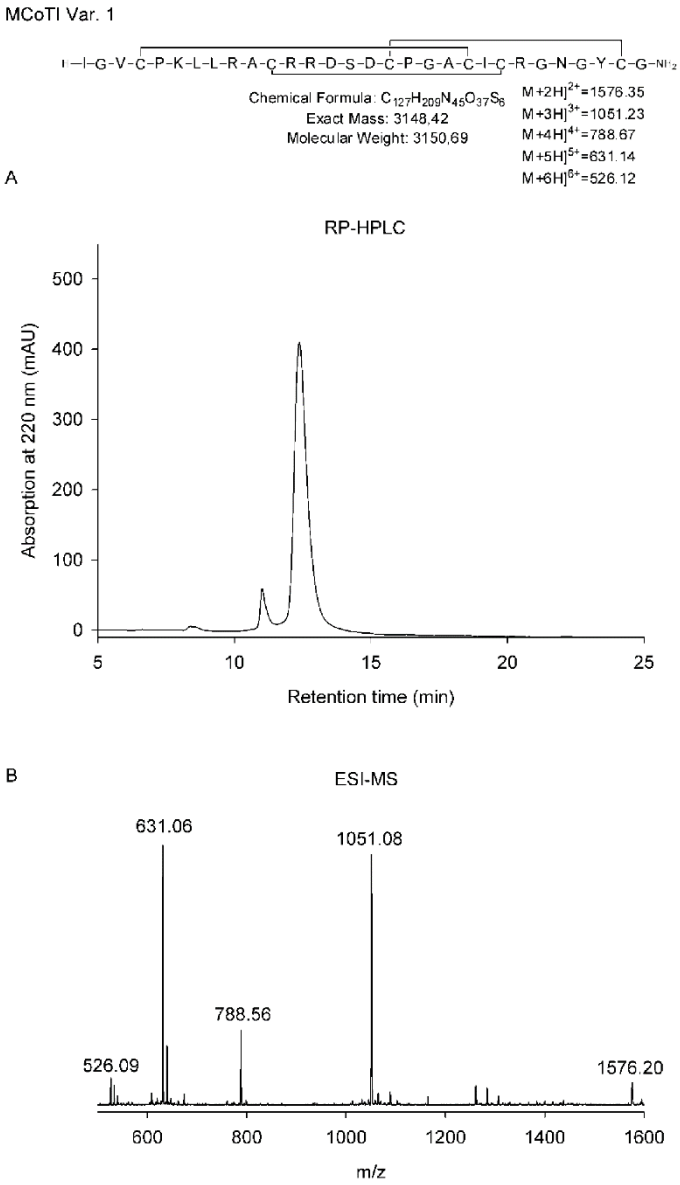


Figure S8: HPLC and MS analysis of MCoTI Var. 1. (**A**) HPLC trace (10 to 60 % B over 20 min) at 220 nm. (**B**) ESI-MS of peptide-containing fraction.

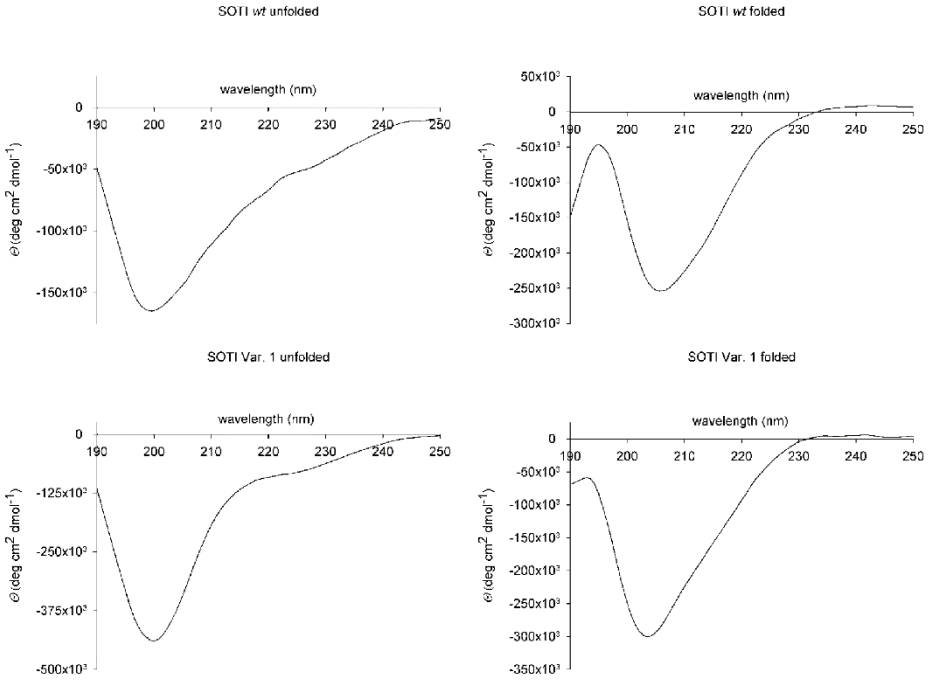


Figure S9: CD spectroscopy of the reduced (unfolded) and oxidized (folded) variants of SOTI *wt* and SOTI Var. 1. Smoothed with the 'smooth' function of Sigma Plot 11.

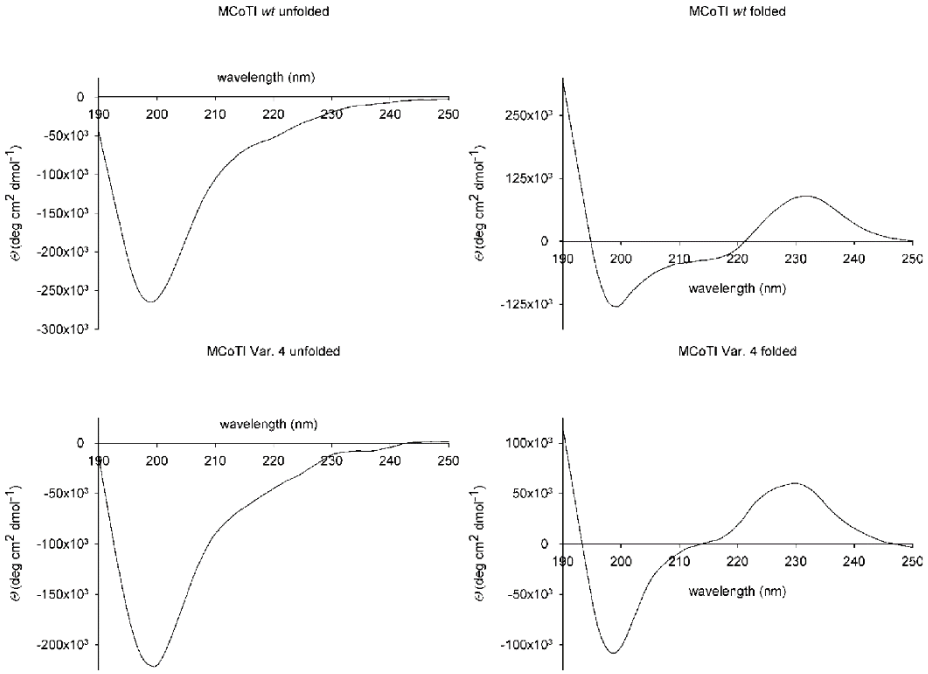


Figure S10: CD spectroscopy of the reduced (unfolded) and oxidized (folded) variants of MCoTI *wt* and MCoTI Var. 4. Smoothed with the 'smooth' function of Sigma Plot 11.

References

1. Heitz A, Hernandez JF, Gagnon J, Hong TT, Pham TT, et al. (2001) Solution structure of the squash trypsin inhibitor MCoTI-II. A new family for cyclic knottins. Biochemistry 40: 7973-7983.
2. Glotzbach B, Schmelz S, Reinwarth M, Christmann A, Heinz DW, et al. (2013) Structural characterization of Spinacia oleracea trypsin inhibitor III (SOTI-III). Acta Crystallogr D Biol Crystallogr 69: 114-120.
3. Hernandez JF, Gagnon J, Chiche L, Nguyen TM, Andrieu JP, et al. (2000) Squash trypsin inhibitors from Momordica cochinchinensis exhibit an atypical macrocyclic structure. Biochemistry 39: 5722-5730.
4. Steinmetzer T, Schweinitz A, Sturzebecher A, Donnecke D, Uhland K, et al. (2006) Secondary amides of sulfonylated 3-amidinophenylalanine. New potent and selective inhibitors of matriptase. J Med Chem 49: 4116-4126.
5. Corpet F, (1988) Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res, 16, 10881–10890.
6. Reinwarth M, Glotzbach B, Tomaszowski M, Fabritz S, Avrutina O, et al. (2013) Oxidative folding of peptides with cystine-knot architectures: kinetic studies and optimization of folding conditions. Chembiochem 14: 137-146.

4.3 Fragmentation follows Structure: Top-Down Mass Spectrometry Elucidates the Topology of Engineered Cystine-knot Miniproteins

Electronic Supporting Information

Fragmentation follows structure: Top-down mass spectrometry elucidates the topology of engineered cystine-knot miniproteins from the MCoTI family

Michael Reinwarth,[1] Olga Avrutina,[1] Sebastian Fabritz*[2] and Harald Kolmar*[1]

1 Clemens-Schöpf Institute of Organic Chemistry and Biochemistry, TechnischeUniversität Darmstadt, Petersenstr. 22, 64287 Darmstadt, Germany.
2AB Sciex Germany GmbH, Landwehrstrasse 54, 64293 Darmstadt, Germany.

Table of contents:

Experimental procedure for mass spectrometric measurements

MS full scans experiments: For the present study the 4000 QTRAP® LC/MS/MS and the 6500 QTRAP® LC/MS/MS systems (AB Sciex Germany GmbH) were used. Peptide stock solutions were prepared by dissolving 0.5 mg of the respective peptide in 1 mL of a mixture of 20 % methanol (Fluka Analytical, LC-MS CHROMASOLV) and 80 % water (Merck KGaA, LC-MS LiChrosolv) with 0.2 % formic acid (Fluka Analytical, analytic additive). The stock solutions were diluted to achieve $1e^6$ counts per second (cps) maximal signal intensity for full scan (m/z 500- 2000, 200 Da/s, unit resolution) experiments. The analyte solutions were infused (10 µL/min) into the masspectrometer using a syringe pump equipped with a 1 mL Hamilton syringe (internal diameter 4.61 mm). The peptide ions were generated using the positive mode electrospray ionization (ESI +). Source and gas parameters are provided in Table S-1.

Table S1: Applied source and gas parameters.

	4000 QTRAP®	**6500 QTRAP®**
Controll Software	Analyst 1.5.2	Analyst 1.6.2
Curtain gas (Cur)	1.38 bar (20 psi)	1.72 bar (25 psi)
Ionization Voltage (IS)	5000 V	5500 V
Temperature (TEM)	Room temp.	Room temp
Sprayer gas (GS1)	1.03 bar (15 psi)	0.89 bar (13 psi)
Heater gas (GS2)	0 bar	0 bar
Declustering Potential (DP)	80 V or ramped*	60 V or ramped*
Entrance Potential (EP)	10 V	10 V

*The declustering potential was semi-automatically ramped and the value resulting in a maximum overall intensity was chosen for the experiment

Enhanced resolution experiments: Enhanced resolution experiments were performed to verify peptide charge states. The parameters resulting from the optimization of full scan experiments were used. However, the DP was semi-automatically re-optimized to yield an maximum intensity for the ion of interest. The first quadruple mass filter enabled ions ±15 Da of the target mass to pass into the linear ion trap. The optimal ion trap fill time was automatically determined (dynamic fill time, DFT) based on a total ion current target (TIC) of $0.5\times1e^7$ cps (QT4000®) or $1\times1e^7$ cps (QT6500®). To ensure an optimal resolution the ion trap scan was performed with a speed of 250 Da/s. A minimum of 100 scans were accumulated.

MS/MS experiments: As a starting point, the parameters resulting from the optimization of full scan experiments were used. However, the DP was semi-automatically optimized to achieve maximum intensity of the parent ion of interest. The collision energy was set by default to 50 V. If not appropriate the parameter was adjusted in 10 V steps or semi-automatically optimized *via* ramping. The optimal ion trap fill time for enhanced product ion scans

was automatically determined (dynamic fill time, DFT) based on a total ion current (TIC) target of $2\times1e^7$ cps (QT4000®) or $10\times1e^7$ cps (QT6500®). Up to 500 scans were accumulated using maximal scans speeds of 4000 Da/s (QT4000®) or 10000 Da/s (QT6500®).

MS^3 Experiments: For MS^3 experiments the presetting from MS/MS measurements were used, while resonance excitation energy (AF2) was semi-automatically optimized (using 'ramping' functionality of Analyst software). For mapping of all MS^3 fragments the smallest AF2 value allowing for full fragmentation of the parent fragment-ion was chosen. Regarding low-intensity fragments, AF2 energy was adjusted to achieve maximized intensities of those specific ions. Excitation times of 25 or 100 ms were used. The Q3 entry barrier was set to 8V.

ESI-MS spectra of *1*, *2* and *3*.

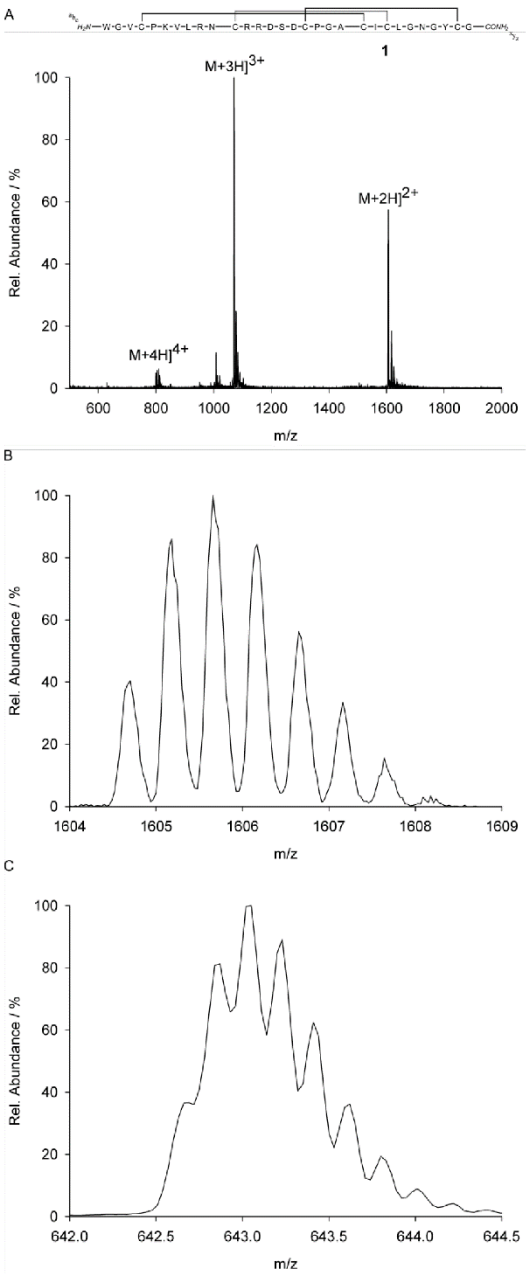


Figure S1: Initial spectra of **1**. (**A**) MS^1 of **1**. (**B**) Zoom-In on $M+2H]^{2+}$. (**C**) Zoom-In on $M+5H]^{5+}$.

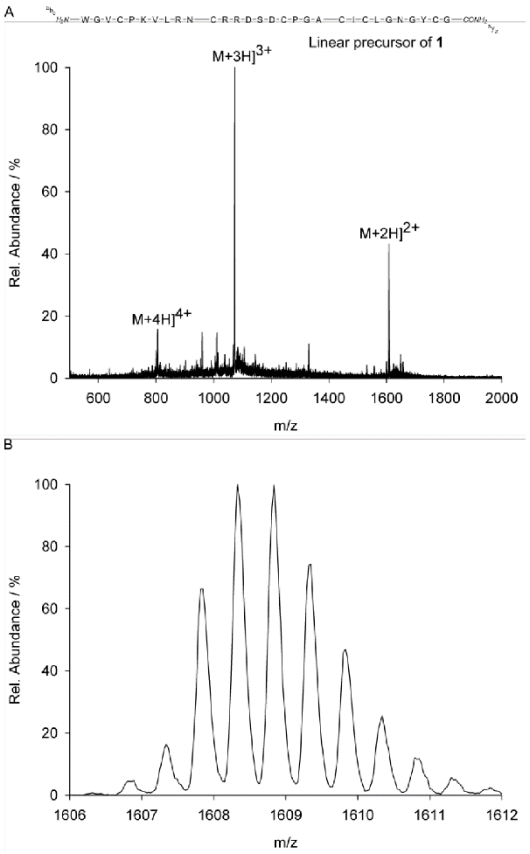


Figure S2: Initial spectra of unfolded precursor of **1**. (**A**) MS^1 of unfolded precursor of **1**. (**B**) Zoom-In on $M+2H]^{2+}$.

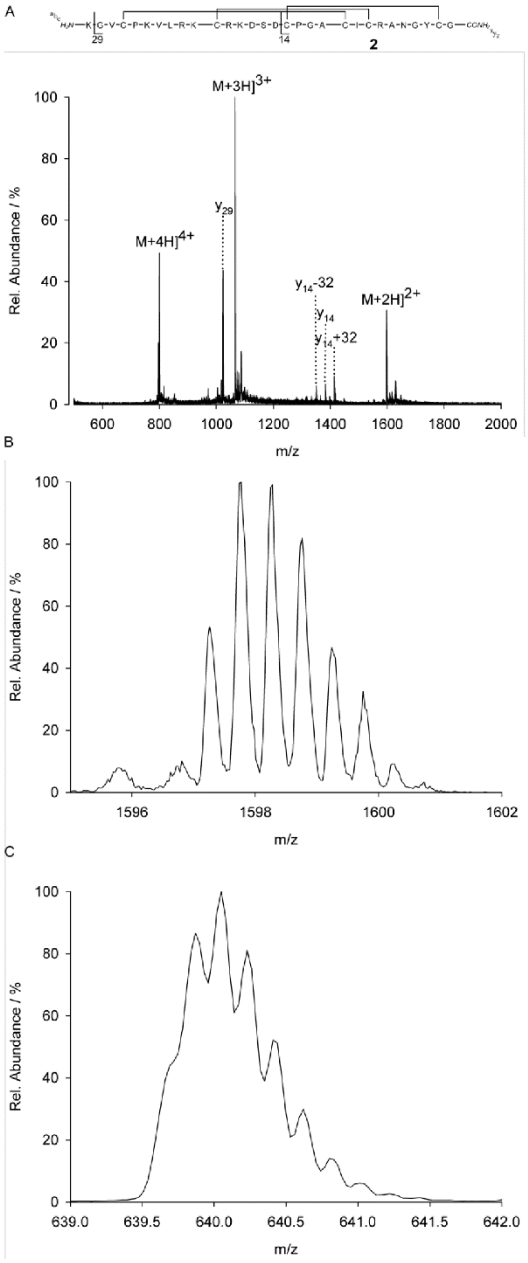


Figure S3: Initial spectra of **2**. (**A**) MS^1 of **2**. (**B**) Zoom-In on $M+2H]^{2+}$. (**C**) Zoom-In on $M+5H]^{5+}$.

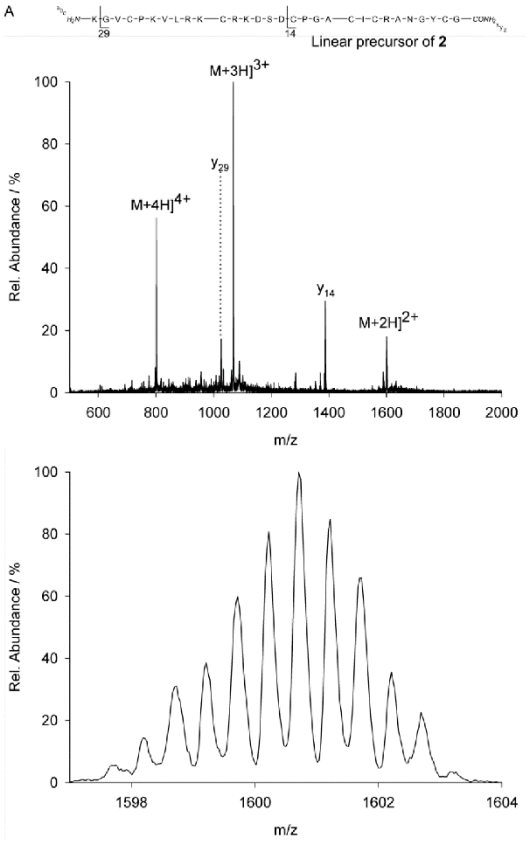


Figure S4: Initial spectra of unfolded precursor of **2**. (**A**) MS^1 of unfolded precursor of **2**. (**B**) Zoom-In on $M+2H]^{2+}$.

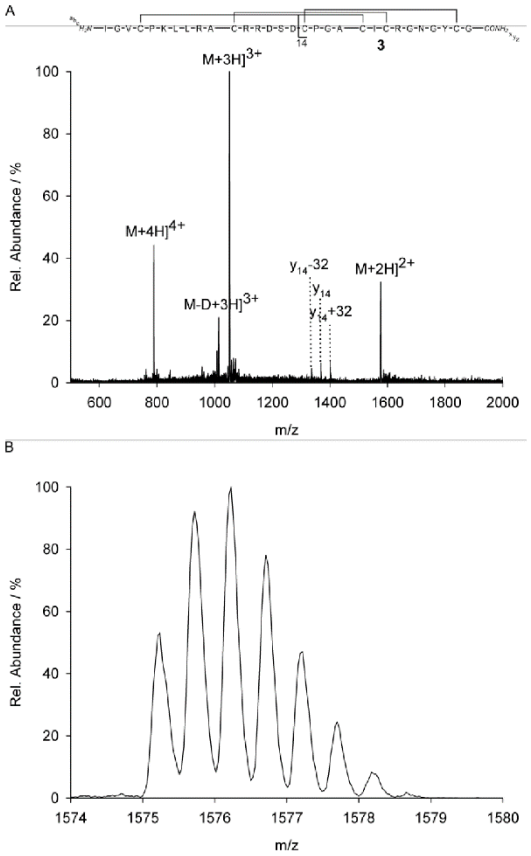


Figure S5: Initial spectra of **3**. (**A**) MS^1 of **3**. (**B**) Zoom-In on $M+2H]^{2+}$.

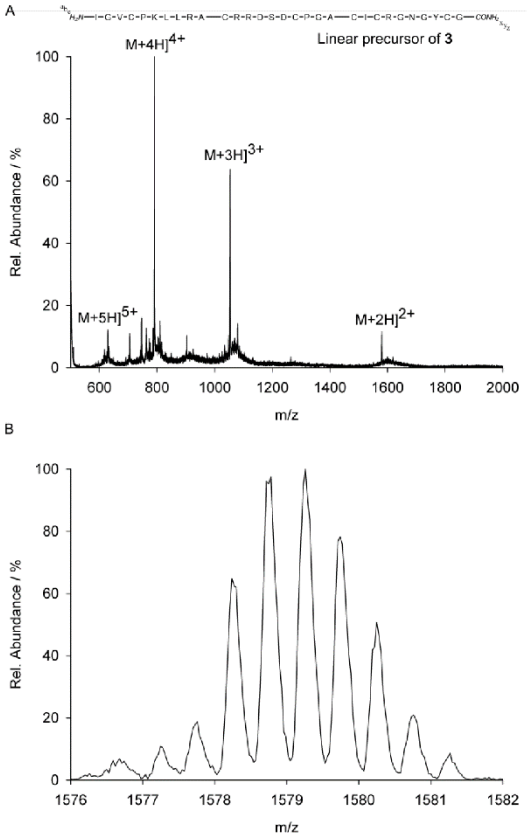


Figure S6: Initial spectra of unfolded precursor of **3** (**A**) MS^1 of unfolded precursor of **3**. (**B**) Zoom-In on $M+2H]^{2+}$.

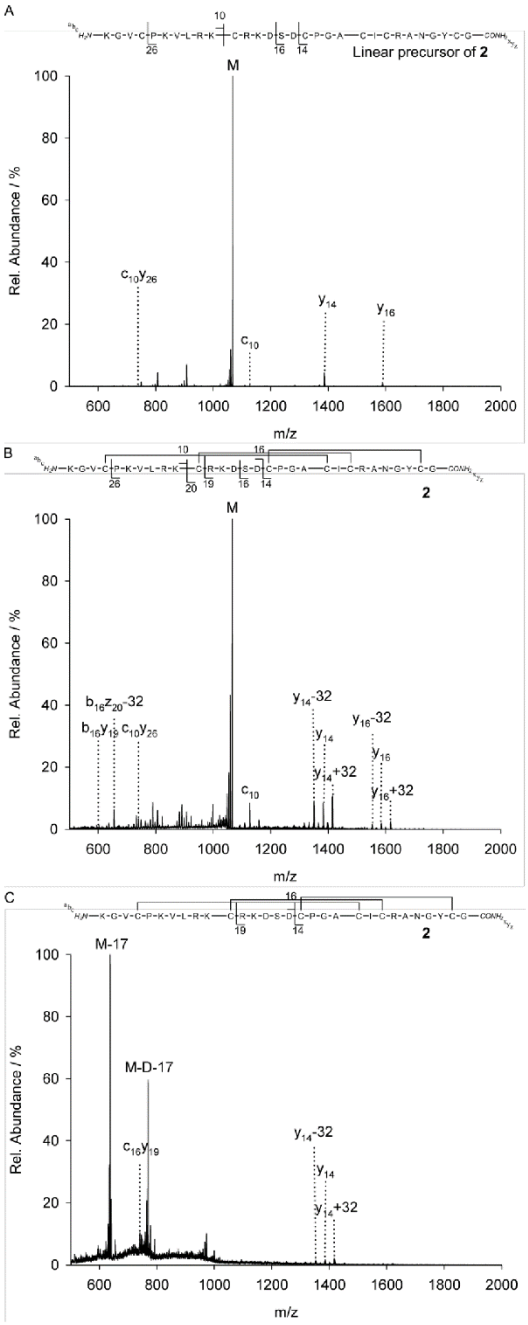


Figure S7: MS^2 of MCoTI peptide **2**. (**A**) CID of the triply charged ion of reduced precursor of peptide **2**. (**B**) CID of the triply charged ion of folded miniprotein **2**. (**C**) CID of the fivefoldly charged ion of folded miniprotein **2**.

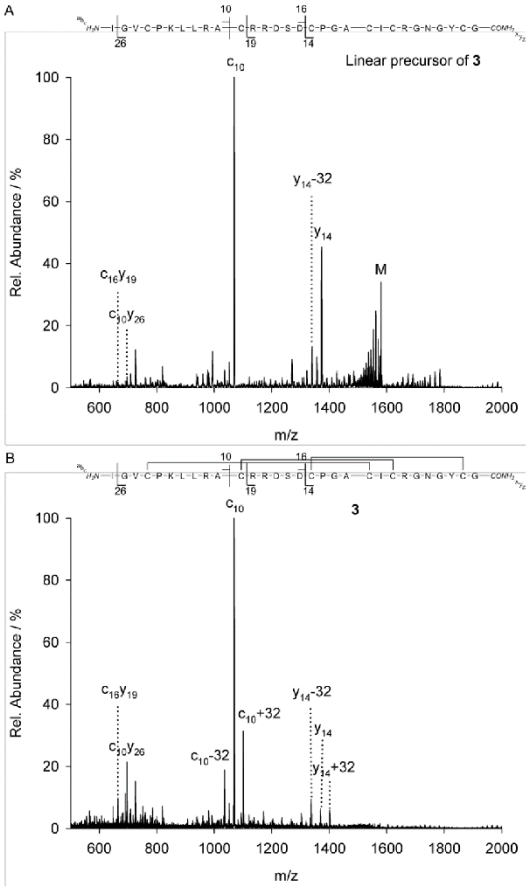


Figure S8: MS^2 of MCoTI peptide **3**. (**A**) CID of the triply charged ion of reduced precursor of peptide **3**. (**B**) CID of the triply charged ion of folded miniprotein **3**.

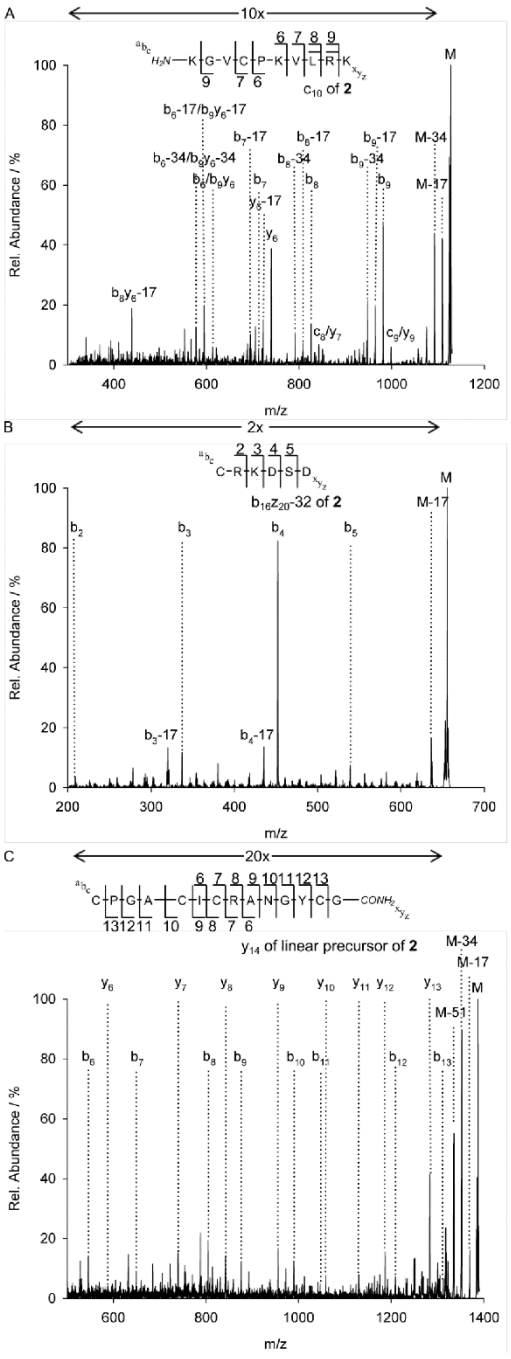


Figure S9: MS^3 of major fragments of MCoTI peptide **2**. (**A**) MS^3 of c_{10} ion. (**B**) MS^3 of $b_{16}z_{20}$ ion. (**C**) MS^3 of y_{14} ion. Arrows above the spectra indicate intensity amplifications.

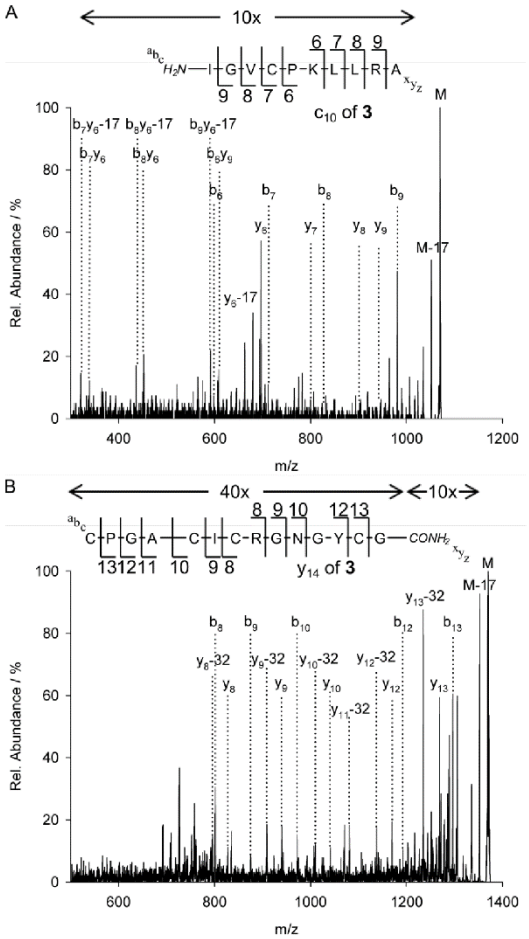


Figure S10: MS^3 of major fragments of MCoTI peptide **3**. (**A**) MS^3 of c_{10} ion. (**B**) MS^3 of y_{14} ion. Arrows above the spectra indicate intensity amplifications.

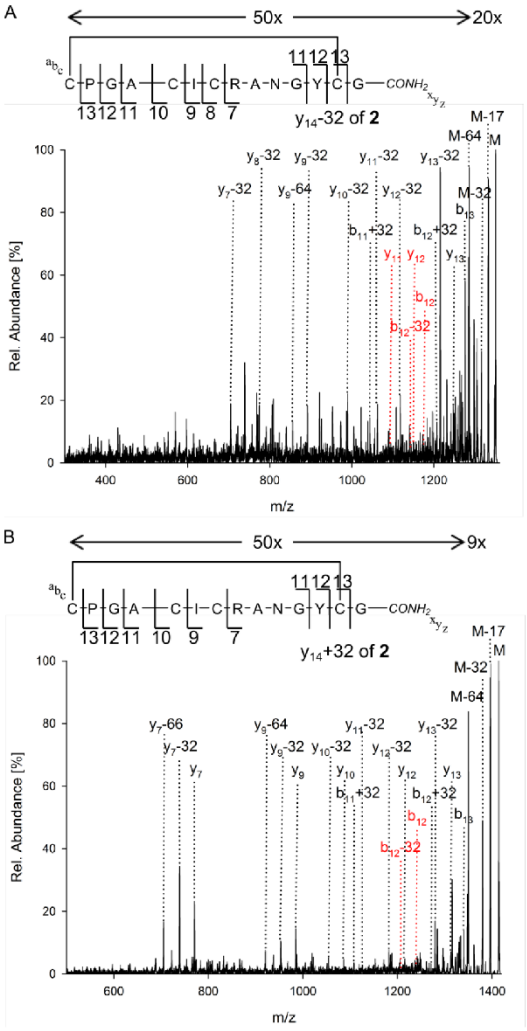


Figure S11: MS^3 of y_{14}-32 (**A**) and y_{14}+32 (**B**) for the combinatorial interpretation of **2**. In red are inexistent peaks to provide evidence on the respective Ptc or Dha cleavage. Arrows above the spectra indicate intensity amplifications.

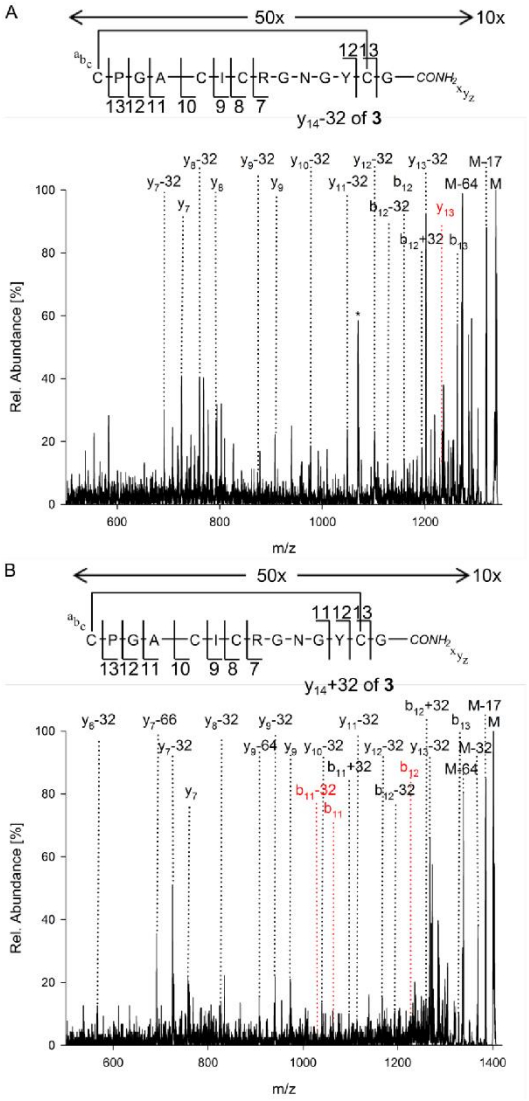


Figure S12: MS^3 of y_{14}-32 (**A**) and y_{14}+32 (**B**) for the combinatorial interpretation of **3**. In red are inexistent peaks to provide evidence on the respective Ptc or Dha cleavage. Asterisk indicates fragment from different parent ion with identical mass. Arrows above the spectra indicate intensity amplifications.

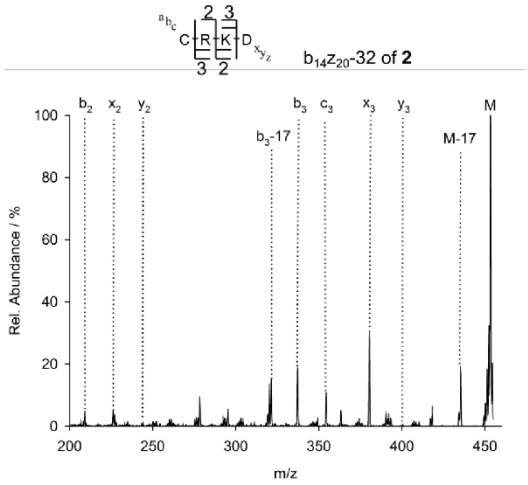


Figure S13: MS^3 of $b_{14}z_{20}$-32 fragment of MCoTI peptide **2**.

Michael Reinwarth

Steubenstr. 53

63225 Langen

20. Mai 2016

Erklärung

Ich erkläre hiermit an Eides statt noch keinen Promotionsversuch unternommen zu haben.

Michael Reinwarth

Michael Reinwarth

Steubenstr. 53

63225 Langen

20. Mai 2016

Erklärung

Ich erkläre hiermit, dass ich meine Dissertation selbständig und nur mit den angegebenen Hilfsmitteln angefertigt habe.

Michael Reinwarth

Lebenslauf

Michael Reinwarth

Persönliche Angaben

Geburtsdatum , -ort	04.05.1981, Langen
Nationalität	deutsch
Familienstand	verheiratet, zwei Kinder

Beruflicher Werdegang

seit 12/2013	Arbeit als *Field Service Engineer* bei *Agilent Technologies*

Promotion

10/2009 bis 11/2013	Promotion im Arbeitskreis von Prof. Dr. Kolmar an der TU Darmstadt

Studium

10/2001 bis 09/2009	**Studium der Chemie** **Abschluss: Diplom Ingenieur Chemie** Technische Universität Darmstadt Fachrichtung Biochemie, Titel der Diplomarbeit: *„Towards Directed Oligomerisation of HIV Entry Inhibitors“*
03/2005 bis 03/2007	Unterbrechung des Studiums (Elternzeit)

Schule und Wehrdienst

07/2000 bis 04/2001	Materialgerätewart beim ABC Abwehrbataillon 750 in Bruchsal
bis 06/2000	Abschluss: **Abitur**
	Dreieichgymnasium Langen

Sprachen

Englisch	Verhandlungssicher in Wort und Schrift

Publikationen

2012 Reinwarth, M.; Nasu, D.; Kolmar, H.; Avrutina, O., Chemical synthesis, backbone cyclization and oxidative folding of cystine-knot peptides: promising scaffolds for applications in drug design. *Molecules* **2012,** 17, (11), 12533-52.

2013 Reinwarth, M.; Glotzbach, B.; Tomaszowski, M.; Fabritz, S.; Avrutina, O.; Kolmar, H., Oxidative folding of peptides with cystine-knot architectures: kinetic studies and optimization of folding conditions. *Chembiochem* **2013,** 14, (1), 137-46.

2013 Glotzbach, B.; Reinwarth, M.; Weber, N.; Fabritz, S.; Tomaszowski, M.; Fittler, H.; Christmann, A.; Avrutina, O.; Kolmar, H., Combinatorial Optimization of Cystine-Knot Peptides towards High-Affinity Inhibitors of Human Matriptase-1. *PLoS One* **2013,** 8, (10), e76956.

2014 Reinwarth, M.; Avrutina, O.; Fabritz, S.; Kolmar, H., Fragmentation follows structure: top-down mass spectrometry elucidates the topology of engineered cystine-knot miniproteins. *PLoS One* **2014,** 9, (10), e108626.

2011 Empting, M.; Avrutina, O.; Meusinger, R.; Fabritz, S.; Reinwarth, M.; Biesalski, M.; Voigt, S.; Buntkowsky, G.; Kolmar, H., "Triazole bridge": disulfide-bond replacement by ruthenium-catalyzed formation of 1,5-disubstituted 1,2,3-triazoles. *Angew Chem Int Ed Engl* **2011,** 50, (22), 5207-11.

2012 Fabritz, S.; Horner, S.; Konning, D.; Empting, M.; Reinwarth, M.; Dietz, C.; Glotzbach, B.; Frauendorf, H.; Kolmar, H.; Avrutina, O., From pico to nano: biofunctionalization of cube-octameric silsesquioxanes by peptides and miniproteins. *Org Biomol Chem* **2012,** 10, (31), 6287-93.

2013 Glotzbach, B.; Schmelz, S.; Reinwarth, M.; Christmann, A.; Heinz, D. W.; Kolmar, H., Structural characterization of Spinacia oleracea trypsin inhibitor III (SOTI-III). *Acta Crystallogr D Biol Crystallogr* **2013,** 69, (Pt 1), 114-20.

2015 Maass, F.; Wustehube-Lausch, J.; Dickgieser, S.; Valldorf, B.; Reinwarth, M.; Schmoldt, H. U.; Daneschdar, M.; Avrutina, O.; Sahin, U.; Kolmar, H., Cystine-knot peptides targeting cancer-relevant human cytotoxic T lymphocyte-associated antigen 4 (CTLA-4). *J Pept Sci* **2015,** 21, (8), 651-60.

Sonstige Interessen

Fußball (aktiv und als Jugendtrainer), Lesen, Computer, Familie und Freunde

Langen, den 20. Mai 2016

Michael Reinwarth